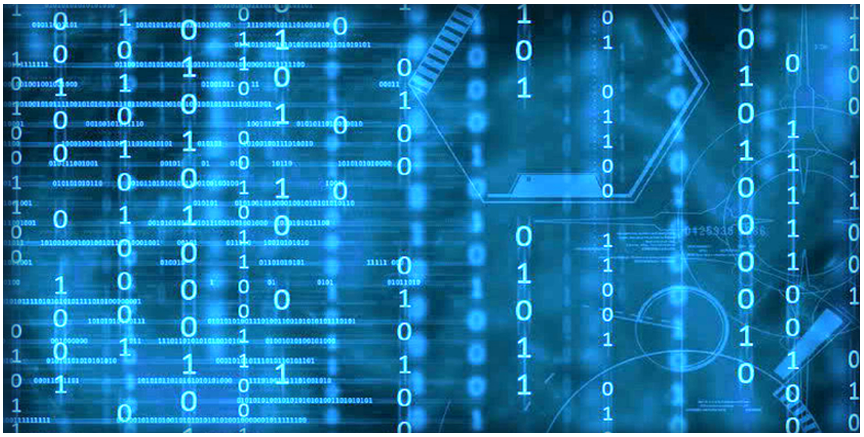
IJACSA

WHERE WISDOM SHARES

International Journal of Advanced Computer Science and Applications

Volume 3 Issue 12

December 2012



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

SAI

www.ijacsa.thesai.org

IJACSA

WHERE WISDOM SHARES

INTERNATIONAL JOURNAL OF

ADVANCED COMPUTER SCIENCE AND APPLICATIONS

SAI

THE SCIENCE AND INFORMATION ORGANIZATION

www.thesai.org | info@thesai.org

Editorial Preface

From the Desk of Managing Editor…

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

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

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

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

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

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

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

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

Thank you for Sharing Wisdom!

Managing Editor
IJACSA
Volume 3 Issue 12 December 2012
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

Reviewer Board Members

CONTENTS

A new approach towards the self-adaptability of Service-Oriented Architectures to the context based on workflow

Faîçal Felhi
Department of Computer Sciences
High Institute of Management, BESTMOD Laboratory
Tunis, Tunisia

Jalel Akaichi
Department of Computer Sciences
High Institute of Management, BESTMOD Laboratory
Tunis, Tunisia

***Abstract*— Distributed information systems are needed to be autonomous, heterogeneous and adaptable to the context. This is the reason why they resort Web services based on SOA Based on the advanced technology of SOA. These technologies can evolve in a dynamic environment in a well-defined context and according to events automatically, such as time, temperature, location, authentication, etc... This is what we call self-adaptability to context. In this paper, we are interested in improving the different needs of this criterion and we propose a new approach towards a self-adaptability of SOA to the context based on workflow and showing the feasibility of this approach by integration the workflow under a platform and test this integration by a case study.**

Keywords- SOA; Webservices; self-adaptability; ubiquitous system; workflow.

I. INTRODUCTION

System Information (SI) must meet some specific constraints surrounding context adaptation in the case of ubiquitous computing [22]. This is particularly the case in areas of industry relating to many domains such as RFID (Radio Frequency Identification) [19][42], e-Learning ect... For this case, considerable approaches related to adaptability with different modes of implementation such as: AOP (Aspect Oriented Programming)[10]. This aspect used by various platforms on the goal to adapt the Web service (WS) [30] to the context dynamic changes of environment. Web services, like any other middleware technologies, aim to provide mechanisms to bridge heterogeneous platforms, allowing data to flow across various programs. The WS technology looks very similar to what most middleware technologies looks like. The emergence of Web services as a model for integrating heterogeneous Web information has opened up new possibilities of interaction and adaptability to context when offered more potential for interoperability. However, from a set of requirements on SOA (Service Oriented Architecture in English) [6], and to provide self adaptation to the context of Web services, we need to integrate more generic connector that takes into account all ambient or distant events.

Based on technology platforms to adapt of Web services to the context, and also gain benefit from the advantages of Web services. These platforms deal with "simple" adaptation to functional and technical exchange purpose it is by no means clustering adaptation to ambient context year. These systems must be used in different contexts depending on the environment of the user profile and the terminal to use. One of the major problems of such systems relates to the context adaptation. It is important that applications adapt to their surroundings [23]. The adaptation software can take many forms, and we refer to system adaptable when a user can interact with the system and, through them, modify and customize. An adaptive system identifies a situation and adapts. Activation of this system changes can be caused by human intervention or a number of observations.

The main goal of context execution is to allow the application to manage the external situations that affect its quality of service seen by the user. Thus, the application must be adaptable to disappearances and appearances of devices on the network, for example, or all sorts of technical failures, malicious or unusual expense. Our goal is to study the adaptability with a presentation of platforms to adapt their operating principles, and trying to improve the SOA to empower the Web services to be self adaptive to the contexts. In our paper we propose an approach for self adaptation of SOA to the context.

The rest of this paper is organized as follows: In Section 2, we review previous research on adaptability of Web services as well as other related issues. Section 3 proposes the needs of self adaptability of SOA to the context. In section 4 we present our approach for self adaptability of SOA to the context and illustrated our work by feasibility and a case study. Finally, we summarize our work and discuss future research in Section 5.

II. RELATED WORKS

Based on aspects, AspectWerkz [12], supporting weaving at loading, offers various means of expression of cut points. It offers different methods to specialize or customize the middleware according to the code on the works. Furthermore, it supports different various methods of expression (XML and Java). DynamicTAO [8] have a particularity that can represent different aspects of the ORB [2] in the form of strategy design pattern. The system configuration is offer via a file (specifies the different strategies applied by the ORB before launching the system). Based on our work, we conclude that this platform suffers like all conventional middleware for the inability to reconfigure the ORB during execution time. DynamicTAO, address aspects such as: competition, safety and supervision, and can provide a set of interfaces allowing users to reconfigure its structure.

Each WS has an Interface Definition Language, namely WSDL (Web Service Description Language) [31], that is responsible for the message payload, itself described with the equally famous protocol SOAP (Object Access Protocol) [32], while data structures are explained by XML (eXtended Markup Language) [33]. Very often, WS are stored in UDDI (Universal Description Discovery and Integration) [34] registry. Based on aspects and Web services, Charfi and al. approach [1] propose a framework that provides support for middleware BPEL (Business Process Execution Language) [37] engines. The authors apply the concepts of deployment descriptor and container for the Web service composition. Ferraz Tomaz and al. approach [24] proposed a tool for weaving aspects for a simple adaptability of the Web services, implementing aspects of the services as loosely coupled, where aspects are woven dynamically. In this approach, aspects are themselves Web services, thus they are independent of languages and platforms. This approach provides a mechanism, based on the AOP for Web services, to dynamically adapt to different policies of use. This approach has two major limitations related to its architecture and its implementation. These limitations are: the dependence of the runtime architecture of Web services and dependence on aspects of language.

Mehdi Ben Hmida approach [18] extended the solution proposed by [24] to specify BPEL processes adaptable, that is to say, the adaptability of complex services. This specification allows the generation of customers that adapt to dynamic changes on the server side. Changes made to the service can lead to further exchanges of information between client and server that were not initially planned during implementation of the client application. These new interactions normally cause runtime errors of the customer. This approach is based on process algebras to dynamically generate customers. Process algebras manage the interactions between a BPEL process and its customers, this by specifying formally the interaction protocol (abstract BPEL) and automatically generating a client who is successfully communicating with the service. This approach overcomes the limitations in the dynamic modification of a process that can lead to a change in the pattern of interaction with the client and will fit the client and server parts. Hence the need to extend the semantic aspects and Web services, which resulted in the ASW (Aspect Service Weaver). Aspects are themselves loosely coupled Web services, they are independent of languages and platforms, but, this approach has limitations. Adaptation to context is not taken into account, that is to say, if an event occurred during a search on a Web service, this approach does not take into account this event.

In the other approaches we find those based on context adaptation. The ambient computing encourages the proliferation of associated devices. A key aspect of the ambient computing is its invisibility. Users perceive the features but do not see the devices that provide these features. Adaptability and evolution of software in these devices becomes an asset to their condition of use. From studies by [4] we can summarize the descriptions of the following platforms. Aura [5] is a context-sensitive middleware that enables the design of mobile applications. Aura's goal is to provide each user with a set of implicit processing services and information. ExORB [21] demonstrates the ability of a middleware to support its configurability, the possibility of putting it-to-day improvement. ExORB examine in particular middleware for mobile phones. Such devices require a middleware on which it is possible to configure new software, to improve the software already built without manual intervention by the end user. ExORB purposed of contributing to the construction of middleware services for strengthening the main features of configurability, the ability to update and improvement. DoAmI (Domain-Specific Ambient Intelligence) [17] deals with the dynamic aggregation of distributed services. Its offers an architectural model that relies on a service-oriented middleware for integration and activation of services at runtime.

CORTEX (Co-operating Real-time sentient objects: architecture and Experimental Evaluation) [9] aims to build a middleware component-based applications to accommodate influenced by the external environment, particularly in the transportation field. CORTEX needs and capabilities of local services for local decision making. The system participates in a cooperative global system. Thus, it has a system for collecting real-time environment. In addition, local systems can be equipped with additional features such as consideration of traffic lights or a mechanism to allow pedestrians (presence sensor, obstacle avoidance). CORTEX defines objects aware. They are moving objects that behave independently and are responsible for interactions with the physical environment. These behaviors are based on sensor inputs and the internal state of each object. To address the problems of coordination, control, adaptability and scalability, CORTEX provides a first programming paradigm using objects aware. The discovery mechanism is not well detailed by the authors and does not measure the scalability of the architecture. In addition, it lacks the tools to do a self adaptation.

WComp [14][15][16] represent the implementation of experimental models presented in the work of RAINBOW (research team of the I3S laboratory, hosted by University of Nice - Sophia Antipolis). This is a platform for lightweight components for service composition SLCA (Service Lightweight Component Architecture) which enables the design of ambient computing applications by assembling software components, orchestrating access to services through infrastructure devices from ambient. WComp supports protocols such as UPnP (Universal Plug and Play) [38] and Web services, allowing components through the proxy to interact with them. WComp is a prototyping “development” environment for context-aware applications. The WComp Architecture is organized around Containers and Designers paradigms. The purpose of the Containers is to take into account system services required by Components of an assembly during runtime: instantiation, destruction of software Components and Connectors. The purpose of the Designers allows configuring assemblies of through Containers. To promote adaptation to context WComp uses Aspect [30] Assembly paradigm. Aspect Assemblies can either be selected by a user or fired by a context adaptation process. It uses a weaver that allows adding and or suppressing components. A container includes a set of (Beans) components and each bean has properties, input methods that use received input information, and output Methods to send to another bean, for instance, output information. Aspect Assemblies allow defining links between Beans by using input and output information. WComp uses UPnP (Plug and Play) technology to detect locally whether the device is active or not and to define input methods and sent events for each component. With this architecture WComp allows: i) managing devices heterogeneity and dynamic discovering by using UPnP, ii) events driven interactions with devices, iii) managing dynamic devices connection and disconnection (dynamic re configuration on run time) in infrastructure.

III. Adaptability Of Soa To The Context And Needs Of Self Adaptability

A. *Adaptability of SOA to the context*

The SOA offer great flexibility that is a great ability to functional and technical changes. Moreover, this type of architecture is most often used as Web services support, which provide the flexibility and interoperability expected, that is the ability to communicate between heterogeneous systems. When the SOA is based on Web services, is referred to as WSOA for (Web Services Oriented Architecture). The application in such information systems that incorporate SOA need to communicate across the exchange software (middleware or platforms). These middleware are the source of our work. It is on them that will think the same expectations in terms of flexibility, interoperability and adaptability.

The adaptability of a system [13] is the software mechanisms that achieve system changes. It is these mechanisms that modify the behavior of the system while preserving the properties of the system. The self-adaptability means the power to dynamically modify the behavior of a system in response to internal and external events. If the user has the possibility to adapt the interaction of these preferences, the interface is said to be configurable and adaptable. If the system is able to adapt his behavior to the needs (capabilities and preferences of the current user) during the interaction, with its capacities of perception and interpretation of the interaction and its context, the interface is called adaptive. The adaptivity of a system is how the system adapts. It consists of strategies to trigger changes in the system based on incentives.

The main goal of context [36] execution is to allow the application to manage the external situations that affect its quality of service seen by the user. Therefore, the application must be adaptable to disappearances and appearances of devices on the network, for example, or all sorts of technical failures, malicious or unusual expense. The context is not directly involved in the application, but it sets the execution environment of the application that is a subset of the software infrastructure. It is in fact available resources of the system at a given [26].

B. *Needs of self adaptability*

The adaptation of service-oriented architectures is at the heart of building new applications.

A self-adaptable architecture is that it takes into account the business logic and HMI (Human Machine Interface). A workflow engine allows you to run a process defined elsewhere in the tool design process that accompanies it. This execution is sequential and is completely unconditional, conditional and is based on a set of rules defining the conditions of connection.

The workflow tools allow the definition of rules more or less sophisticated but still generally quite simple and few: Boolean operators, data fields of the process values entered by operators, properties of any documents involved in the process, etc…

The workflow engine can then connect to the rules engine to "know" what option to take during the course of a process. The rules engine also allows users to expose simple interfaces for generating these rules.

Following the research and the state of the art described above, lead us to determine the list of needs:

- An SOA for its flexibility, interoperability with Web services.

- Use of aspects for adaptation to changes that do not provide Web services.

- The use of such powerful concepts: reflexivity and / or MOP (Meta Object Protocol) [25] to gain flexibility and performance, or dynamic reconfiguration.

- Adaptation to the prevailing situation with mechanisms and platforms to capture the events and process them in real time, or in a near real time, by invoking the right service.

- Consideration of transverse functions such as security, log management, etc...

- Consideration of management rules intrinsically.

- The management of processes and data.

- Interoperable intelligent connectors that can be applied to the platform regardless of its technology.

IV. Self-Adaptability Of Soa To The Context Based On Workflow

In this section we present the functional and technical architecture of our new approach for the self adaptability to the context of SOA based on the needs that we have already cited.

A. *Functional architecture WSOA with WWF*

Fig. 1 shows our needs in terms of self adaptability of service oriented architecture to the context. We present in the following the functionality provided by our system:

- Decentralized and reconfiguration: our architecture is based on objects or components to make the dynamic reconfiguration of components using more advanced mechanisms. It qualifies the distribution of applications across multiple servers and not the increase in service levels. is a distributed architecture whose purpose is to deliver services to their audience and they will be accessible from any types of clients.

- Event and communication management: The events sent by the external environment are adapted to the context. This architecture must include an inference engine, which specifies the behavior of applications in a given context and uses the execution model events-conditions-actions. Management communication, use the mechanism based on events that are created dynamically during system operation. The analysis of the communication is based on modeling the communication events in composite services for ambient spaces.

- Weaving and generation: Weaving is the process that takes as input a set of aspects and a base application, and outputs an application whose behavior and structure are extended by aspects. Generating code corresponding to the component assembly, manipulation of the graphical representation of the application, generation of the executable code corresponding to the application. The weaving can Occur at compile time (Modifying the compiler), load time (Modifying the class loader) or runtime (Modifying the interpreter).

- Work flow and rules management: The workflow engine can then connect to the rules engine to "know" what option to take during the course of a process. The rules engine also allows users to expose simple interfaces for generating these rules.

- Composition and flow orchestration: enables the design of ambient computing applications by assembling software components and orchestration of access to services by devices from the ambient infrastructure.

- Security and administration: Offered by this system in treating the business logic from the workflow and rules.

- Discovery: Contextual resource discovery is the use of context data to discover other resources within the same context.

- Invocation ambient and distant services: The invocation of distant and ambient services is also permitted by this architecture using technologies dedicated to each type of invocation.

B. *Technical architecture WSOA with WWF*

This architecture (Fig. 2) allows the structuring of technical capabilities and infrastructure in our new approach to SOA.

In this architecture, we present the different functionality by connectors well integrated within our system. What is specific in our approach, we propose to integrate connectors rules Engine that communicates with a workflow engine in our framework. In this rules engine we need to define the rules that manage the data flow to finally produce events providing services to the customer. The data management shall be provided from a component that specifies the service to send it to another component by assembling them that allows the management process, event management and orchestration services. Web services are invoked by remote proxy with their specific WSDL URL, as well as ambient services using specific technologies such as UPnP.

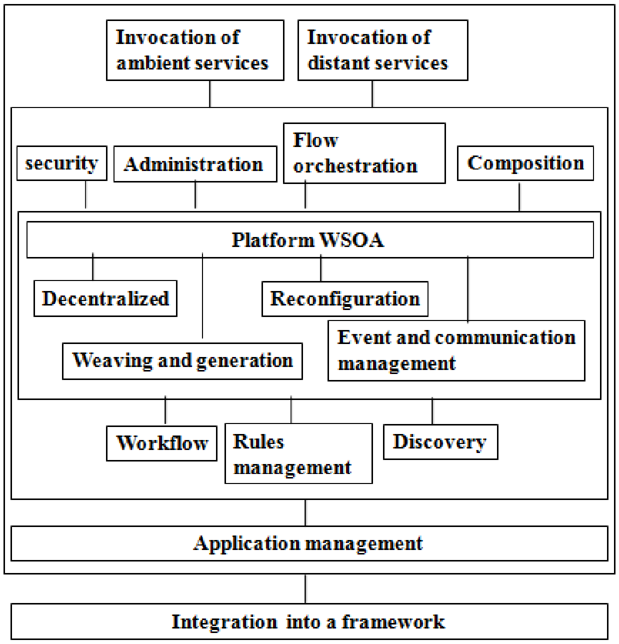


Figure 1. Functional Architecture.

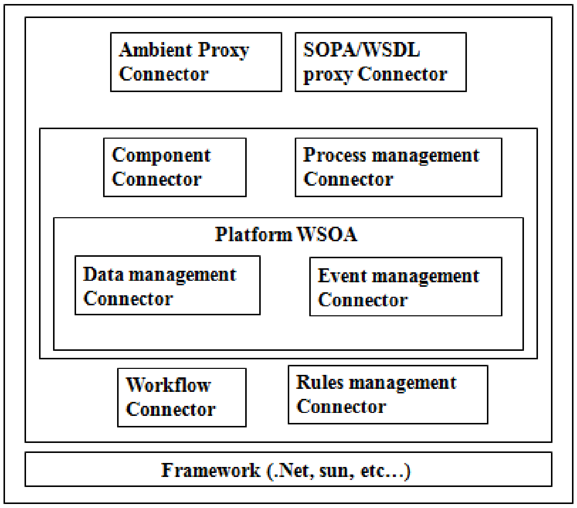

Ambient Proxy Connector
SOPA/WSDL proxy Connector
Component Connector
Process management Connector
Platform WSOA
Data management Connector
Event management Connector
Workflow Connector
Rules management Connector
Framework (.Net, sun, etc…)

Figure 2. Technical Architecture.

C. *Feasibility*

In our research [26] [27] [28] we studied some platforms and in particular WComp and CORTEX, whose our main goal is to propose a self-adaptive SOA to the context. These studies have led to several implementations and case studies in different fields, such as, e-learning, smart house, RFID, etc... Result of these studies, we presented in our work [7] a proposal to an self-adaptable SOA when we've built a rules engine within WComp which can offer management rules that deal with business logic. Business logic can help in the development and optimization of these assemblies separating the events produced by the components defined in a WComp application.

Under WComp we have integrated a rule engine that can provide management rules that deal with business logic.

The rules engine can communicate with a workflow engine, which helps optimize and evolution of these assemblies separating the events produced by the components defined in an application WComp.WWF (Windows Workflow Foundation) [40][41] is a framework that allows users to create flow systems or user applications written for Windows Vista, Windows XP and Windows Server 2003 family. WWF can solve simple scenarios such as displaying user interface controls based on user input or complex scenarios encountered by large companies, such as order processing and inventory control. WWF treat the flux activation in business applications, the flow of pages the user interface, the flow of paper, user flows, mixed flows for applications based on services, and flow-driven rules business.

Our solution represented in Fig. 3 is based on the WWF under .Net that can solve simple scenarios such as displaying user interface controls based on user input or complex scenarios encountered by large enterprises. This integration solves simple scenarios such as displaying user interface controls based on user input or complex scenarios encountered by large enterprises.

In this architecture, except for the different needs initially used by WComp (service invocation ambient and remote data orchestration ...) describe above in the related word section, we integrated connectors rules engine that communicates with a workflow engine in framework .Net. In this rules engine we need to define the rules that manage the data flow to finally produce events providing services to the customer. The information shall be provided from a component that specifies the service to send it to another component by assembling them in a container through the language of Aspect of Assembly.

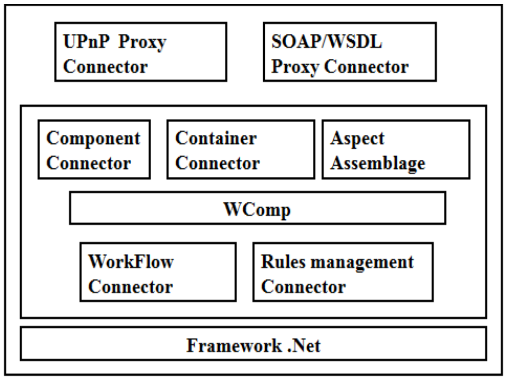


Figure 3. Technical Architecture WComp integrating workflow.

D. Case Study

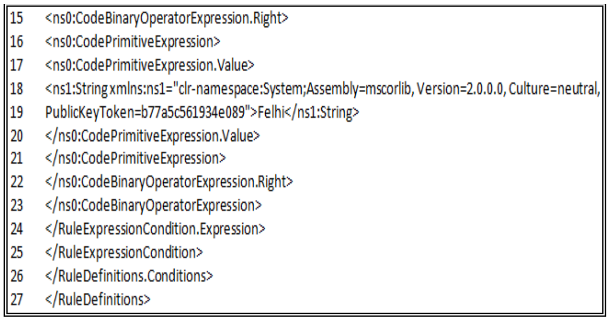
1) Description

We chose to take a sample case study of authentication. This authentication is supposed to capture with a RFID sensor using UPnP technology in WComp. This authentication can, thereafter, to monitor such access to a well determined. In our case, and to simulate this case study, we created a man-machine interface that captures a user authentication; this authentication is then verified based on business rules defined in XML, to finally show a message validation or invalidation of the value set as authentication.

2) Implementation

In a first step, we built the container WComp under sharpdevelop1.0.2a [39]. This container contains all components necessary to make the bean test authentication with an assembly between them. In a second step, we imported the container WComp sharpdevelop3.2 under a project to integrate the flows and rules. We chose to show the code (Fig. 4) of the rule that deals with the flow and data entered by the user. This code is in XML form.

In this block we defined as an example the rule that gives the exact value of authenticating a user. As shown in line 19, the exact value is “Felhi”. This value is normally detected by RFID and simulated in this example by entering a “TexField”, and for example displayed on a screen that is simulated by a “label”.



```
15    <ns0:CodeBinaryOperatorExpression.Right>
16    <ns0:CodePrimitiveExpression>
17    <ns0:CodePrimitiveExpression.Value>
18    <ns1:String xmlns:ns1="clr-namespace:System;Assembly=mscorlib, Version=2.0.0.0, Culture=neutral,
19    PublicKeyToken=b77a5c561934e089">Felhi</ns1:String>
20    </ns0:CodePrimitiveExpression.Value>
21    </ns0:CodePrimitiveExpression>
22    </ns0:CodeBinaryOperatorExpression.Right>
23    </ns0:CodeBinaryOperatorExpression>
24    </RuleExpressionCondition.Expression>
25    </RuleExpressionCondition>
26    </RuleDefinitions.Conditions>
27    </RuleDefinitions>
```

Figure 4. Setting value.

V. CONCLUSIONS

In this paper we presented our new approach for a self-adaptability of service-oriented architectures to the context based on workflow by presenting the functional and technical architecture. In our solution we have given different features in terms of the needs of self-adaptability offered by the integration of workflow, which allows the management rules and a kind of security and administration of Web services. This solution which can offer management rules that deal with business logic. Business logic can help in the development and optimization of these assemblies separating the events produced by the components of Web services.

We have shown the feasibility of this approach by making use WComp platform and we integrated WWF workflow and we tested this integration through a case study.

This solution is a first step towards our problem of defining a service-oriented middleware self-adaptable to the context. We want in our future work added other connectors generic for the more needs of self adaptability.

REFERENCES

[1] A. Charfi and M. Mezini. “Aspect-Oriented Web Service Composition with AO4BPEL”, (2004), 2nd European Conference on Web Services (ECOWS) Publisher, vol. 3250 of LNCS, Springer, pp. 168-182.

[2] C. Schmidt Douglas and C. Cleeland. “Applying Patterns to Develop Extensible ORB MIddelware”. (1999), IEEE Communications Publisher, Magazine Special Issue on Design Patterns Publisher.

[3] C. Ullrich; K. Borau; H. Luo; X. Tan; L. Shen and R. Shen. Why Web 2.0 is Good for Learning and for Research: Principles and Prototypes. Proceedings of the 17th International World Wide Web Conference, ACM, 2008.

[4] D. Cheung-Foo-Wo, M. Riveill, and J-Y Tigli, “Adaptation dynamique par tissage d’aspects d’assemblage”, (2009), Thesis I3S, NICE-SOPHIA ANTIPOLIS.

[5] D. Garlan, D. P. Siewiorek, A. Smailagic, and P. Steenkiste, “Aura : Toward distraction free pervasive computing”, (2002), IEEE Pervasive Computing Publisher.

[6] F. Curbera, R. Khalaf, N. Mukhi, “Quality of Service in SOA Environments”. An Overview and Research Agenda (Quality of Service in SOA-Umgebungen). it - Information Technology 50(2): 99-107, 2008.

[7] F. Felhi and J. Akaichi, “Adaptation of Web services to the context based on workflow: Approach for self-adaptation of service-oriented architectures to the context”, (2012) International Journal of Web & Semantic Technology (IJWesT) Vol.3, No.4, October 2012 Publisher.

[8] F. Kon, M. Roman, P. Liu, J. Mao, T. Yamane, L. Magalhaes, and R. Campbell, “Monitoring, security and dynamic configuration with the dynamictao reflective orb”, (2000), Middleware’2000 Publisher, New York, USA.

[9] G. Biegel and V. Cahill, “A framework for developing mobile, context-aware applications”, (2004), 2nd IEEE Conference on Pervasive Computing and Communication, pp.361–365.

[10] G. Kiczales, J. Lamping, C. Maeda, and C. Lopes, “Aspect-oriented programming”, (1997), Proceedings European Conference on Object-Oriented Programming (ECOOP’97) , volume 1241, pp 220–242. Springer- Verlag, Berlin, Heidelberg, and New York.

[11] G. Reese, “Cloud Application Architectures: Building Applications and Infrastructure in the Cloud”, (2009), O’Reilly Publisher.

[12] J. Bonér , “AspectWerkz: Dynamic AOP for Java”, (2004) AOSD Publisher.

[13] J-Y Tigli, S. Lavirotte, D. Cheung-Foo-Wo., (2003) “Mobilité et Enseignement à Distance” (special issue: TICE Seminar) International Journal of Information Science and Communication (ISDM) Publisher, number 10, ISSN 1265-499X.

[14] J.-Y. Tigli, S. Lavirotte, G. Rey, V. Hourdin, and M. Riveill, “Lightweight Service Oriented Architecture for Pervasive Computing”, (2009), IJCSI International Journal of Computer Science Issues, Vol. 4, No. 1, , ISSN (Online): 1694-0784, ISSN (Print): 1694-0814.

[15] J.-Y. Tigli, S. Lavirotte, G. Rey, V. Hourdin, and M. Riveill, “Context-aware Authorization in Highly Dynamic Environments”, (2009), IJCSI International Journal of Computer Science Issues, Vol. 4, No. 1, , ISSN (Online): 1694-0784, ISSN (Print): 1694-0814.

[16] J.-Y. Tigli, S. Lavirotte, G. Rey, V. Hourdin, D. Cheung-Foo-Wo, E. Callegari, and M. Riveill. “WComp Middleware for Ubiquitous Computing: Aspects and Composite Event-based Web Services”, (2009), Annals of Telecommunications, volume 64, n° 3-4, pp 197. ISSN 0003-4347.

[17] M. Anastasopoulos, H. Klus, J. Koch, D. Niebuhr, and E. Werkman, “DoAmI – a middleware platform facilitating re-configuration in ubiquitous systems”, (2006), System Support for Ubiquitous Computing Workshop. At the 8th Annual Conference on Ubiquitous Computing (Ubicomp) Publisher.

[18] M. Ben Hmida, R. F. Tomaz, and V. Monfort, “Applying AOP concepts to increase Web services flexibility”, (2006), Journal of Digital Information Management (JDIM) Publisher.

[19] M. Hughes, “RFID tags for ambient intelligence: present solutions and future challenges”, Joint sOc-EUSAI conference, Grenoble, 12-14 Octobre 2005.

[20] M. Miller , “Cloud Computing: Web-Based Applications That Change the Way You Work and Collaborate Online”, (2008), QUE Publisher.

[21] M. Roman and N. Islam, “Dynamically programmable and reconfigurable middleware services”, (2004), Middleware Publisher, Vol. 3231 in LNCS, pp. 372–396, Springer.

[22] M. Weiser, “Some Computer Science Issues in Ubiquitous Computing”, (1993), Communications of the ACM, vol. 36, no. 7, pp. 75–84.

[23] N. Ferry and S. Lavirotte, (2008) “Adaptation Dynamique d'Applications au contexte”, I3S laboratory, hosted by University of Nice - Sophia Antipolis.

[24] R. Ferraz Tomaz, M. M. Ben Hmida, and V. Monfort, “Concrete Solutions for Web Services Adaptability Using Policies and Aspects”, (2006), JDIM - Journal of Digital Information Management Publisher.

[25] S. Chiba and T. Masuda, “Designing an Extensible Distributed Language with a Meta-Level Architecture”, (1993), ECOOP’93, pp.482-501, LNCS 707, Springer-Verlag, Kaiserslautern, Germany.

[26] S. Lavirotte, D. Lingrand, and J-Y Tigli. , “Définition du contexte : fonctions de coût et méthodes de sélection”, (2005), Proceedings of the 2nd French-speaking conference on Mobility and ubiquity computing, pp. 9-12.

[27] V. Monfort and F. Felhi, “Context Aware Management Platform to Invoke remote or local e Learning Services Application to Navigation and Fishing Simulator”, International Symposium on Ambient Intelligence, (2010), ISAMI'10 Publisher, special volume in Advances in Intelligent and Soft Computing (Springer), Guimarães, Portugal.

[28] V. Monfort, and F. Felhi, « A contextual approach to invoke intelligent house Services: an application to help physically handicapped persons » 1rst International Workshop on Recent Trends in SOA Based Information Systems in conjonction with ICEIS 2010, Funchal Madeira, Portugal, Juin 2010.

[29] [29] V. Monfort, M. Khemaja, N. Ammari, and F. Felhi, « Using SaaS and Cloud computing For "On Demand" E Learning Services : Application to Navigation and Fishing Simulator », short paper in The 10th IEEE International Conference on Advanced Learning Technologies, July 5-7, 2010 Sousse, Tunisia.

[30] Web URL - http://www.w3.org/TR/ws-arch/ (2004).

[31] Web URL - http://www.w3.org/TR/wsdl20/ (2007).

[32] Web URL - http://www.w3.org/TR/SOAP (2007).

[33] Web URL - http://www.w3.org/XML/ (2012).

[34] Web URL - http://www.uddi.org/pubs/uddi_v3.htm (2004).

[35] Web URL - http://javaboutique.internet.com/articles/WSApplications/ (2012).

[36] Web URL - http://www.larousse.fr/dictionnaires/francais, (2010).

[37] Web URL - http://www6.software.ibm.com/software/developer/library/ws-bpel.pdf, (2003).

[38] Web URL - http://www.upnp.org/, (2012).

[39] Web URL - http://community.sharpdevelop.net/, (2011).

[40] Web URL - http://www.bpmbulletin.com/2006/06/21/difference-entre-workflow-et-moteur-de-regle/, 2006.

[41] Web URL -http://www.vdocsoftware.com/vdoc/easysite/InVDOC2010/news/innovation/agilite-regles-metiers, 2010.

[42] Web URL - http://www.rfidfr.org/, 2012.

AUTHORS PROFILE

Faîçal Felhi received the master degree in Intelligent Information System in 2010. He is currently a PhD student in the BESTMOD laboratory in the high institute of management of Tunis. - Tunisia. He is actually teaching in the high institute of Computer and Mathematic of Monastir - Tunisia.

Jalel Akaichi received his PhD in Computer Science from the University of Sciences and Technologies of Lille (France) and then his Habilitation degree from the University of Tunis (Tunisia) where he is currently an Associate Professor in the Computer Science Department. He has published in international journals and conferences, and has served on the program committees of several international conferences and journals. He is currently the Chair of the Master Science in Business Intelligence. He visited and taught in many institutions such as the State University of New York, Worcester Polytechnic Institute, INSA-Lyon, University of Blaise Pascal, University of Lille 1, etc…

Monte Carlo Ray Tracing Based Sensitivity Analysis of the Atmospheric and the Ocean Parameters on Top of the Atmosphere Radiance

Kohei Arai [1]
Graduate School of Science and Engineering
Saga University
Saga City, Japan

***Abstract*—Monte Carlo Ray Tracing: MCRT based sensitivity analysis of the geophysical parameters (the atmosphere and the ocean) on Top of the Atmosphere: TOA radiance in visible to near infrared wavelength regions is conducted. As the results, it is confirmed that the influence due to the atmosphere is greater than that of the ocean. Scattering and absorption due to aerosol particles and molecules in the atmosphere is major contribution followed by water vapor and ozone while scattering due to suspended solid is dominant contribution for the ocean parameters.**

Keywords-Monte Carlo Ray Tracing; radiative transfer; scattering and absorption; geophysical parameters (the atmosphere and the ocean).

I. INTRODUCTION

It is not easy to solve Radiative Transfer Equation: RTE when the RTE includes radiative transfer in the ocean. There are some widely used RTE software codes, 6S, MODTRAN, and the others. These RTE models do not take into account the radiative transfer in the ocean; there are some RTE models for the RTE models for the ocean, though. RTE model proposed here is based on Monte Carlo Ray Tracing: MCRT model [1]. Therefore, it is assumed any materials, particles, molecules, and the others in the ocean.

Considerable geophysical parameters of the atmosphere and the ocean are assumed for investigation of sensitivity of the geophysical parameters on the Top of the Atmosphere: TOA radiance. The following section describes the proposed method for sensitivity analysis including MCRT simulation model. Preliminary simulation results are described together with the simulation results for sensitivity analysis followed by conclusion with some discussions.

II. PROPOSED METHOD

A. *Monte Carlo Ray Tracing Simulation Model*

Illustrative view of the proposed MCRT simulation model is shown in Figure 1. Photon from the sun is input from the top of the atmosphere (the top of the simulation cell). Travel length of the photon is calculated with optical depth of the atmospheric molecule and that of aerosol.

There are two components in the atmosphere; molecule and aerosol particles while three are also two components, water and particles; suspended solid and phytoplankton in the ocean. When the photon meets molecule or aerosol (the meeting probability with molecule and aerosol depends on their optical depth), then the photon scattered in accordance with scattering properties of molecule and aerosol [2].

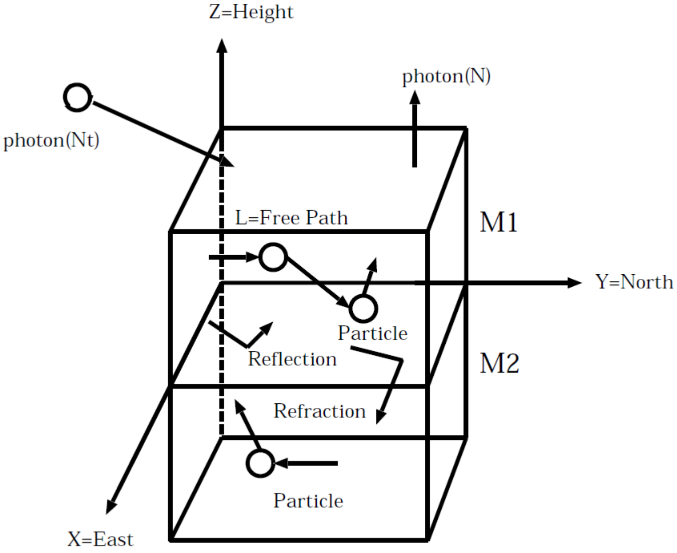


Figure 1. Proposed MCRT simulation model

For simplifying the calculations of the atmospheric influences, it is assumed that the atmosphere containing only molecules and aerosols. Thus the travel length of the photon at once, L is expressed with equation (1).

$$L=L_0 \text{ RND}(i) \quad (1)$$

$$L_0=Z_{max}/\tau \quad (2)$$

where $Z_{max,}$ τ, RND(i) are maximum length, altitude of the atmosphere, optical depth, and i-th random number, respectively. In this equation, τ is optical depth of molecule or aerosol. The photon meets molecule when the random number is greater than τ.

Meanwhile, if the random number is less than τ, then the photon meats aerosol. The photon is scattered at the molecule or aerosol to the direction which is determined with the phase function and with the rest of the travel length of the photon.

Reflection, transpiration, and refraction of the photon at the sea surface are followed by Fresnell law [3] and Snell law [3] as shown in Figure 2.

In three dimensional simulation cells, photon direction has to be changed when the photon meets aerosol particle, molecule, ocean surface, suspended solid, phytoplankton, and so on in accordance with the rotation matrix shown in equation (3) and in Figure 3. In accordance with the number which put at the top left corner, the direction is changed.

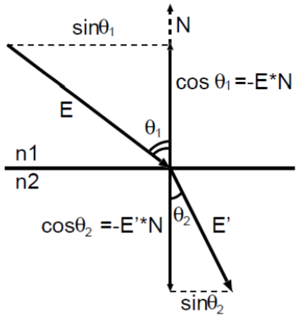
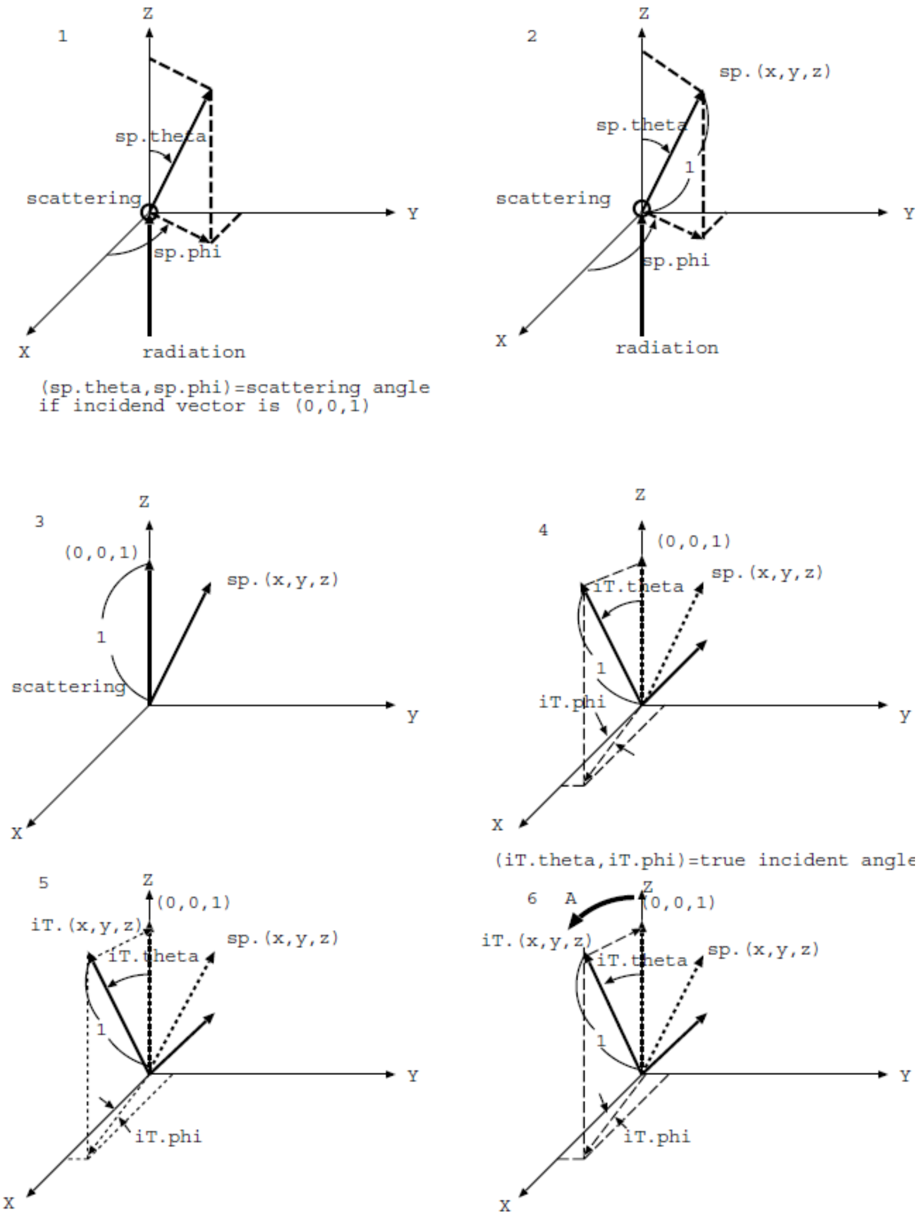


Figure 2. Two dimensional expression of reflection, transpiration, and refraction of the photon at the sea surface

$$\begin{bmatrix} \mathrm{cos(theta)cos(phi)cos(psi)-sin(phi)sin(psi)} & \mathrm{-cos(theta)cos(phi)sin(psi)-sin(phi)cos(psi)} & \mathrm{sin(theta)cos(phi)} \\ \mathrm{cos(theta)sin(phi)cos(psi)+cos(phi)sin(psi)} & \mathrm{-cos(theta)sin(phi)sin(psi)+cos(phi)cos(psi)} & \mathrm{sin(theta)sin(phi)} \\ \mathrm{-sin(theta)cos(psi)} & \mathrm{sin(theta)sin(psi)} & \mathrm{cos(theta)} \end{bmatrix} \quad (3)$$



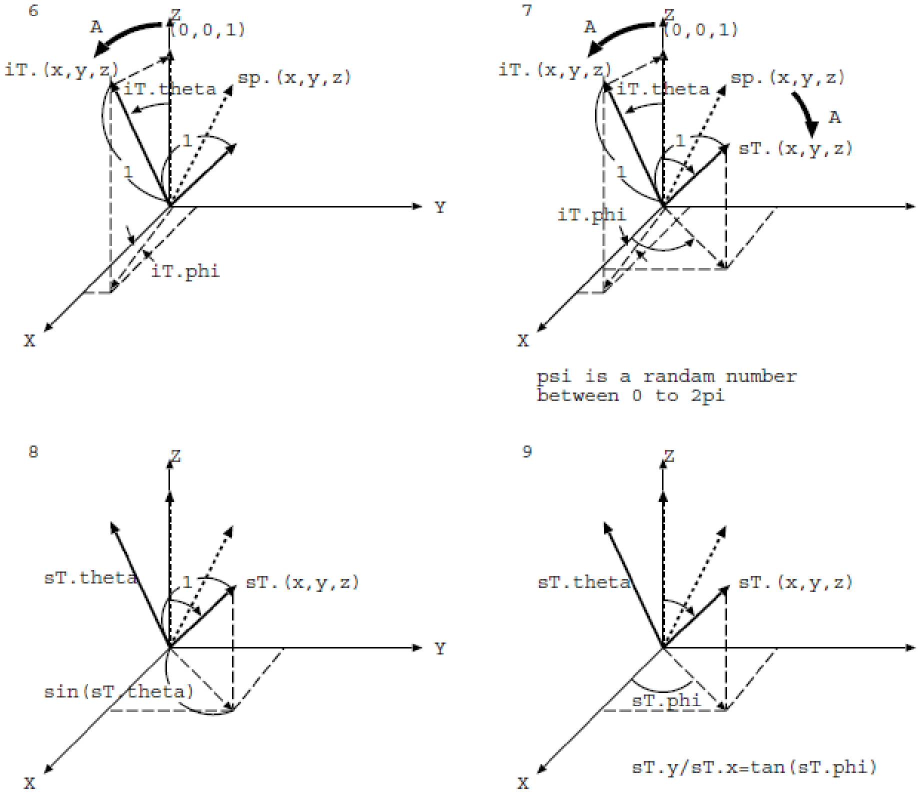


Figure 3. Process flow for photon direction changes when the photon in concern meets aerosol particle, molecule, ocean surface, suspended solid, phytoplankton, and so on

Where phi denote azimuth and psi denote elevation angles. In Figure 3, particle is situated in the center (origin of the three dimensional coordinate system). The phone comes from the bottom and meets with the particle in concern. The scattering angle is expressed as a function of sp.theta and sp.phi is the incident vector is (0,0,1) while the true incident angle is a function of iT.theta and iT.phi in this figure. Then rotation matrix is reduced as shown in equation (3). This process flow is same for reflection, refraction, and scattering.

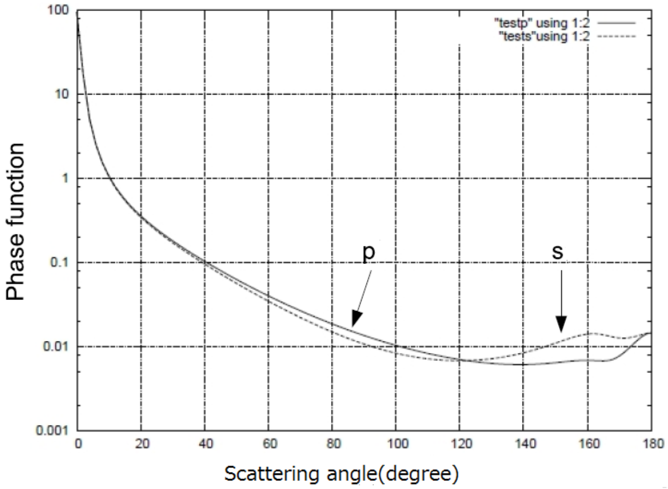
The scattering property is called as phase function. In the visible to near infrared wavelength region, the scattering by molecule is followed by Rayleigh scattering law [3] while that by aerosol is followed by Mie scattering law [3]. Example of phase function of Mie scattering is shown in Figure 4 (a) while that of Rayleigh scattering is shown in Figure 4 (b). In the figure, scattering angle is defined as the angle between incidence and reflected angle from the particle. These phase functions can be calculated with Mie Code in the MODerate resolution atmospheric TRANsmission; MODTRAN[1].

B. Top of the Atmosphere: TOA Radiance Calculation

If the photon reaches on the wall of the simulation cell, the photon disappears at the wall and it appears from the corresponding position on the opposite side wall. Then it travels with the rest of travel length. Eventually, the photons which are reached at the top of the atmosphere are gathered with the Instantaneous Field of View: IFOV of the Visible to Near Infrared Radiometer: VNIR onboard satellite. At sensor radiance, I^+ with direction and IFOV of μ, μ_0 can be calculated with equation (4)

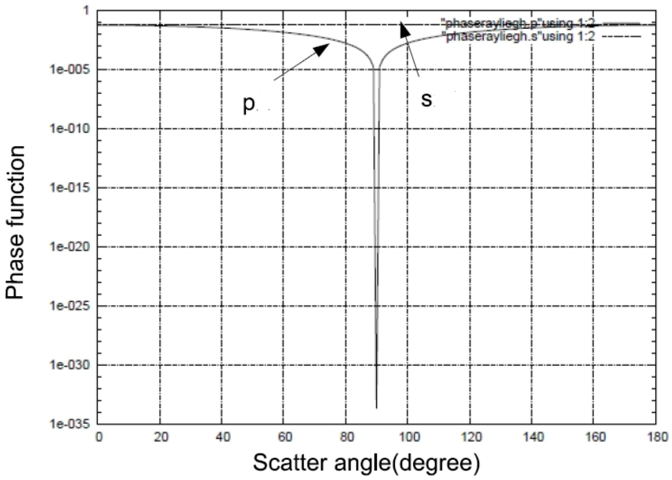
$$I^+(\mu, \mu_0)=I\,N^+(\mu, \mu_0)/N_{total} \qquad (4)$$

where N^+ is the number of photons which are gathered by VNIR, N_{total} denotes the number of photons input to the simulation cell. Also I denotes extraterrestrial irradiance at the top of the atmosphere.



(a)Mie scattering

[1] http://modtran.org/



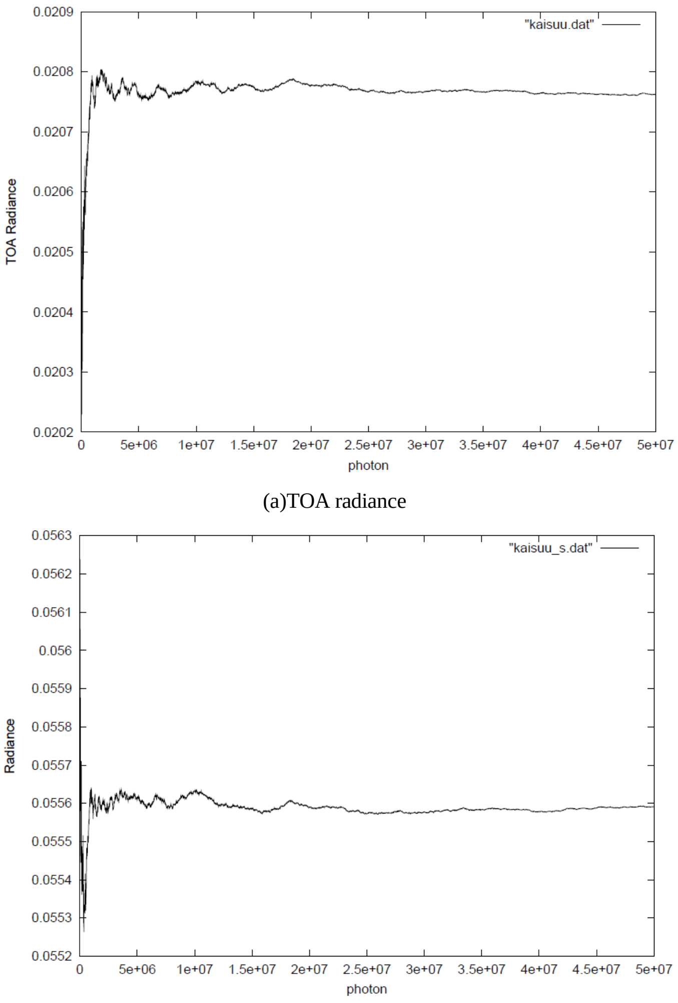
(b)Rayleigh scattering

Figure 4. Phase functions for Mie and Rayleigh scattering

III. EXPERIEMNTS

A. Preliminary Simulations

Preliminary simulation results are shown in Figure 5.



(a)TOA radiance

(b)Water leaving radiance

Figure 5. Examples of the preliminary simulation

In the preliminary simulation study, TOA radiance and water leaving radiance is calculated when the number of photons is changed from 1 to 50 million. The size of simulation cell is defined as 50km by 50km by 50km for the atmosphere while that is set as 50km by 50km by 500m (in depth) for the ocean. Wavelength is set at 500nm while the solar azimuth and zenith angles are set at 90 and 30 degrees. All the parameters required for the simulation are set as follows,

Optical depth of suspended solid: 0.03

Molecule optical depth: 0.25

Aerosol optical depth: 0.3

Single scattering albedo of suspended solid: 0.3

Single scattering albedo of molecule in the ocean: 0.15

As the results, it is found that water leaving radiance is almost twice much greater than that of TOA radiance. Also TOA radiance includes pass-radiance (the photons are scattered in the atmosphere and then come out from the top of the atmosphere without reaching the ocean surface). Therefore, less than 1/10 of small number of photons are come out from the top of the atmosphere from the ocean surface in comparison to the photons which are come out from the atmosphere from the atmosphere. Both TOA and water leaving radiance are saturated at the number of photons is around 30 million. Therefore, the number of photons is set at 30 million for the detailed simulation study on sensitivity analysis.

B. Sensitivity Analysis

Radiance_S and Radiance_T denote water leaving radiance and TOA radiance excluding the contribution due to scattering component in the atmosphere. Meanwhile TOA radiance includes all the contributions from the atmosphere and the ocean. Figure 6 (a) shows these radiances as a function of solar zenith angle. Default parameters for this simulation is as follows,

Optical depth of suspended solid: 0.015

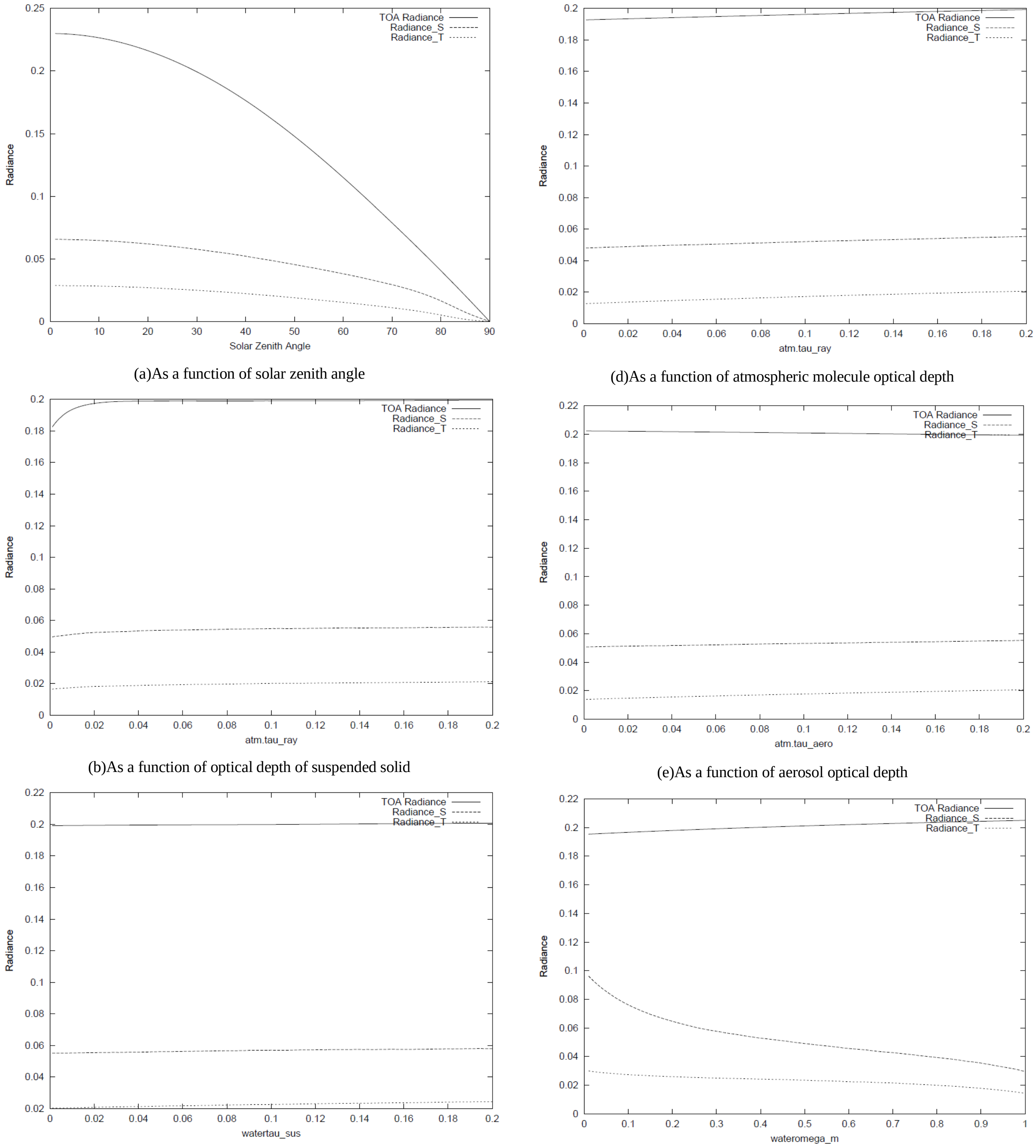
Molecule optical depth: 0.15

Aerosol optical depth: 0.2

Single scattering albedo of suspended solid: 0.6

Single scattering albedo of molecule in the ocean: 0.3

All the parameters are set as default except optical depth of suspended solid. Then the simulation result of TOA radiance, water leaving radiance and TOA radiance excluding scattering component in the atmosphere as a function of optical depth of suspended solid is obtained and is shown in Figure 6 (b). Meanwhile, those radiances as functions of oceanic molecule optical depth, atmospheric molecule optical depth, aerosol optical depth, single scattering albedo of suspended solid, and single scattering albedo of molecule in the ocean are shown in Figure 6 (c), (d), (e), (f), and (g), respectively. From these figures, it is found that TOA radiance is not so sensitive to the suspended solid optical depth, oceanic molecule optical depth, atmospheric molecule optical depth, aerosol optical depth, single scattering albedo of suspended solid and is sensitive to single scattering albedo of molecule in the ocean comparatively.



(a)As a function of solar zenith angle

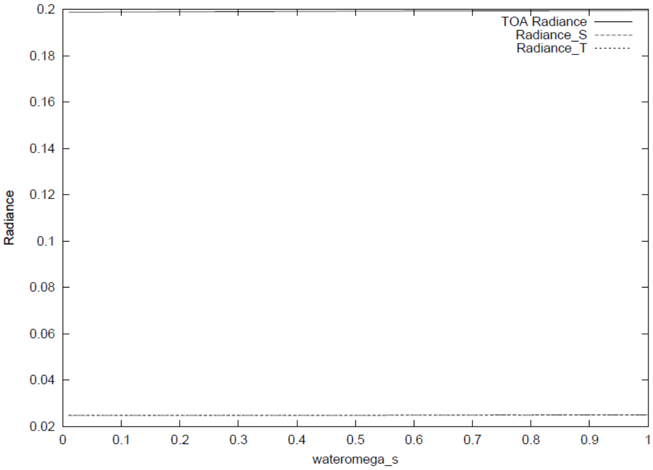
(b)As a function of optical depth of suspended solid

(c)As a function of oceanic molecule (sea water) optical depth

(d)As a function of atmospheric molecule optical depth

(e)As a function of aerosol optical depth

(f)As a function of single scattering albedo of suspended solid



(g)As a function of single scattering albedo of molecule in the ocean (sea water)

Figure 6. TOA radiance, water leaving radiance and TOA radiance excluding the contribution due to scattering in the atmosphere

TOA radiance changes are summarized in Table 1. TOA radiance changes for suspended solid optical depth is greatest followed by sea water optical depth, single scattering albedo of suspended solid, atmospheric molecule optical depth, aerosol optical depth, and single scattering albedo of sea water.

TABLE I. TOA RADIANCE CHANGES

Contribution	TOA_radiance_changes
Oceanic_molecule_optical_depth	0.02813
Suspended_solid_optical_depth	0.1994
Atmospheric_molecule_optical_depth	0.006545
Aerosol_optical_depth	0.003067
Oceanic_molecule_single_scattering_albedo	0.000761
Suspended_solid_single_scattering_albedo	0.009676

C. Bi-directional Reflectance Distribution Function: BRDF

Bi-directional Reflectance Distribution Function: BRDF is estimated with the same default parameters. Solar zenith angle is set at 90 degree. Figure 7 shows the estimated ocean surface reflectance as a function of observation zenith angle. There is sun glint at around 80 degree of observation zenith angle because the solar zenith angle is 90 degree.

IV. CONCLUSION

Monte Carlo Ray Tracing: MCRT based sensitivity analysis of the geophysical parameters (the atmosphere and the ocean) on Top of the Atmosphere: TOA radiance in visible to near infrared wavelength regions is conducted. As the results, it is confirmed that the influence due to the atmosphere is greater than that of the ocean. Scattering and absorption due to aerosol particles and molecules in the atmosphere is major contribution followed by water vapor and ozone while scattering due to suspended solid is dominant contribution for the ocean parameters.

It is also found that the sensitivity on TOA radiance for suspended solid optical depth is greatest followed by sea water optical depth, single scattering albedo of suspended solid, atmospheric molecule optical depth, aerosol optical depth, and single scattering albedo of sea water.

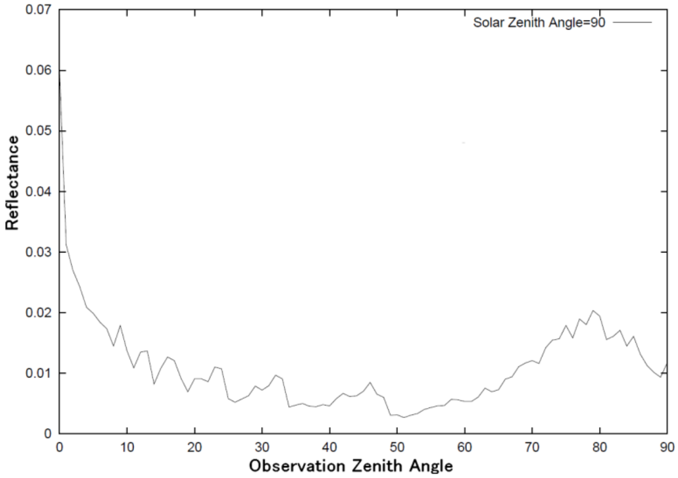


Figure 7. Estimated ocean surface reflectance as a function of observation zenith angle

ACKNOWLEDGMENT

The author would like to thank Dr. Yasunori Terayama and Mr. Keiji Ono for their contributions to conduct simulation studies.

REFERENCES

[1] Kohei Arai, Adjacency effect of layered clouds estimated with Monte-Carlo simulation, Advances in Space Research, Vol.29, No.19, 1807-1812, 2002.

[2] Kohei Arai, Lecture Note for Remote Sensing, Morikita Publishing Inc., (Scattering), 2004.

[3] Kohei Arai, Fundamental Theory for Remote Sensing, Gakujutsu-Tosho Publishing Co., Ltd.,(Lambertian), 2001.

AUTHORS PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science, and Technology of the University of Tokyo from 1974 to 1978 also was with National Space Development Agency of Japan (current JAXA) from 1979 to 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He was appointed professor at Department of Information Science, Saga University in 1990. He was appointed councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was also appointed councilor of Saga University from 2002 and 2003 followed by an executive councilor of the Remote Sensing Society of Japan for 2003 to 2005. He is an adjunct professor of University of Arizona, USA since 1998. He also was appointed vice chairman of the Commission "A" of ICSU/COSPAR in 2008. He wrote 30 books and published 332 journal papers.

Statistical Analysis of the Demographic Ageing Process in the EU Member States, Former Communist Countries

Dorel Săvulea
University of Craiova
Department of Informatics
Craiova, Romania

Nicolae Constantinescu
University of Craiova
Department of Informatics
Craiova, Romania

***Abstract*— The aim of this paper is to make an analysis and a comparative study of the demographic ageing process in the former communist countries which are currently EU member states. Taking into account the complexity of the phenomenon, the study approaches only a part of the indices involved in this process. According to the structure of the population on groups of ages and sexes, the population pyramid is built and in this context we carried out various analyses and comparisons among countries. Further we approach determining demographic factors for the ageing process of the population, as for example the evolution of the young population, the mortality rate and the life expectancy and the changes in the population structure are pointed out with the help of the demographic dependency rate. There have been made a series of statistical correlations and predictions which allowed for a more concrete explanation of the evolution of the demographic ageing phenomenon.**

Keywords- component; statistical analysis; demographic analyse; statistical correlation.

I. Introduction

Demography is considered a social science, with a unitary character, its main objective being the study of population and demographic phenomena. Its basis was set by John Graunt (1620-1674) and it can be defined as the science which studies the population dynamics which comprises the size, structure and demographic age phenomenon as well as its implications.

In the last two decades, the demographic ageing process has increased on a global level which represents an important growth of the aged population segment of the total number of population employed in the socio-economic field. The phenomenon is also found in the EU member states, with special features due to the socio-cultural and economic systems [1].

The population aging combined with the decrease of the birth rate is one of the major problems from an economic, budgetary and social point of view. Some EU countries are developing some programs to fight against this phenomenon especially through reforms of the retirement, health and social care systems.

Due to the lack of such institutional changes and of adequate policies this ageing process will have a strong impact on the economic growth leading to the growth of the public expenses but also of the ones from the infrastructure, education etc.

The debates from the European Council have led to the conclusion that there is a need to manage the effect of the population aging on the economies of the EU countries. Therefore, the Hampton Court Summit from October 2008, the renewed Lisbon Strategy, the new EU development strategy have set the need for a long term coordination of the actions between the member countries in what the retirement, health and care systems are concerned as well as a strategy for the increase of the young population and of the public finances in the context of the Stability and Growth Pact [8]. These debates have shown a margin of manoeuvre of almost 10 years when the labour force continues to grow, and this fact will allow the accomplishment of some specific reforms for an aging society.

This paper seeks to identify, analyse and predict the evolutions of the phenomenon in the former communist countries, now EU member states. It has been conducted a comparative study of the indices involved and the consequences of this process have also been presented. For the statistical data processing we used analysis and data processing software as well as MINITAB and SPSS.

II. Statistical Demographic Implied models

A. *Study on population. Descriptive notions.*

For the analysis of the aging process the definition of some concepts is needed and also their calculation methodology.

For the concept number of population we distinguish two categories: the number of registered population and the calculated number of population.

Definition 1 [11]

a) *the registered number of population* is the number determined after the population survey representing the physical number of population at the moment of registration. With the help of this number one can determine the density of population as well as the indices for the average rhythm of growth of the population;

b) *the calculated number of population* is its number determined with the help go various calculation methods for a date or a given period, for which there are no information;

From the point of view of the dynamic or time series, the number of population might be an *indicator for the moment* (the number of population on the date of the survey or at another moment), but also *interval indicator* (the average number of population in a certain period), for most of the descriptive statistical data which require the use of the number of population as an interval indicator. The calculation formula for the population number is the following:

$$P1 = P0 + (N(0,1) - M(0,1)) + (I(0,1) - E(0,1)) = P0 + \Delta sn + \Delta sm \quad (1)$$

where

P0 şi P1 – population number at two different dates;
N(0,1) – the number of living new born babies in the 0,1 interval;
M(0,1) – the number of deceased in the 0,1 interval;
I(0,1) – the number of immigrants in that population in the 0,1 interval;
E(0,1) – the number of emigrants in that population in the 0,1 interval;
Δsn – the natural growth determined as a difference N(0,1) – M(0,1);
Δsm – the migrating growth determined as a difference I(0,1) – E(0,1).

The two basic demographic characteristics in the analysis of the population structure are the sex and the age. The population structure represents a statistical distribution of a population according to the *qualitative* characteristic *sex* in two sub-populations: masculine and feminine. The population structure according to age represents a statistical distribution of a population according to the *quantitative* characteristic *age*. Usually, for the demographic calculations, population is distributed on the following intervals: 0-4; 5-9; 10-14; …; 95-99; +100.

The most important tool for the analysis of the structure according to age is represented by the so called *population pyramid*, used in the study of the process of *demographic aging of the population*. The pyramid designates a special graphical representation of the population distribution according to age, consisting of two charts, one for the masculine and the other for the feminine population. On the vertical we can see the age of the population, from 0 years old (the pyramid base) to 100 years old (its top), in the left there is the masculine population and in the right the feminine one.

As a consequence of the population decrease due to ageing, the chart becomes a triangle, thus explaining the name of population pyramid.

B. *Population ageing*

The demographic ageing of the population represents a demographic process which relies in the growth of the old population combined with the decrease in the young population; usually the proportion of the adult population remains unchanged for a long time.

The determining factors of the ageing process are the decrease of the birth rate and the mortality.

In this context, it has to be mentioned the fact that the demographic ageing of the population due to the decrease in the birth rate is also known under the name of *ageing at the bottom of the pyramid*, and the one resulted from the decrease of mortality and consequently to the growth of the average life expectancy is called *ageing on the top of the pyramid*.

In order to study the demographic ageing process, according to the International Labour Organization(ILO) and EUROSTAT standard, the population is divided into three groups: 0-14 years old (young population); 15-64 years old (adult population or, in some references, named people of working age); 65+ years old (old population). The share in percentage is calculated according to the formulae:

$$Pt = (P0\text{-}14/Nt) \bullet 100 \quad (2)$$

$$Pv = (P65+/Nt) \bullet 100 \quad (3)$$

where

Pt – percentage share of the young population;
Pv – percentage share of the old population;
Nt – total number of population
P0-14 – total number of population with ages between 0-14 years old,

It is thought that a population where the 65 years old groups and over has less than 7% of the total cannot be regarded as *young* from a demographic point of view. Under the circumstances when the share of the elder people fluctuates between 7-12%, that population is in a *full demographic ageing process*, and when it exceeds 12%, we are talking about a demographic *aged* population.

Due to the fact that the structure according to ages is directly involved in setting the productive potential of the society, it is recommended to set a *dependency ratio*, which measures the pressure exerted by the population from the inactive age groups on the adult population. This ration expresses the number of young and old people from a population below and over the limit of the working age which goes to 100 people which are able to work, belonging to that population.

The *dependency ratio* is determined according to the relation:

$$Rd = [(P0\text{-}14 + P65+)/P15\text{-}64] \bullet 100 \quad (4)$$

The demographic dependency ratio is sometimes divided into the old age dependency ratio (the ration between the 65 years old population and the population between 15-64 years old), written down as Rdv and the young age dependency ratio (the ration between the population below 15 years old and the population between 15-64 years old), written down as Rdc.

$$Rdv = P65+/P15\text{-}64 \bullet 100 \quad (5)$$

$$Rdc = P0\text{-}14/P15\text{-}64 \bullet 100 \quad (6)$$

The structure of the population of the countries might be different according to the fertility, mortality and migration models, from the past and present, which are characteristic for each country, thus distinguishing four general profiles of the population structure, Fig. 1.

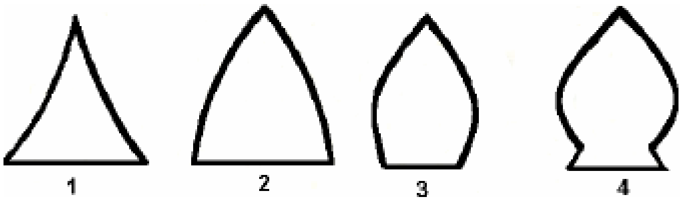


Figure 1. Types of people pyramids

Type 1, also called *circumflex pyramid* is specific for a young population from a demographic point of view, with a high birth rate and mortality.

Type 2, also called *rick* is characteristic for the countries where the ageing process is not advanced and the fertility is relatively high.

Type 3, the *urn pyramid*, represents a population with high demographic ageing symptoms, as a consequence of the significant decrease of fertility, leading to the de-population process.

Type 4, the *clubs pyramid* is characteristic for a population which is getting younger, after a demographic ageing process.

The general mortality rate (RGM) [2,10]**.** It is calculated as a ratio between the number of deceases in a period (usually year) and the average number of population from that year. The formula is:

$$RGM = (D/P)\bullet 1000 \quad (7)$$

where
D – amount of deceases;
P – average number of population;

RGM expresses the number of deceases for 1.000 inhabitants in a random population, in a year.

The mortality rate on ages (RMV). It is calculated as a ratio between the number of deceases at a certain age (age group), in a given period of time (usually a year) and the total number of population of that age (age group). The formula is:

$$RMV= (Dv/Pv)\bullet 1.000 \quad (8)$$

where
Dv –the number of deceases for the young population (age group) v;
Pv – the number of old population of *x* years old or from the *v* age group;

Mortality rate on sexes (RDM, RDF). It is calculated by reporting the number of deceases registered for the male and female persons, in a given period of time, to the number of persons belonging to that sex.

$$RDM = (Dm/Pm) \bullet 1.000 \text{ or } RDF= (Df/Pf) \bullet 1.000 \quad (9)$$

where
Pf, Pm – the number of women and men;
Dm, Df – the number of deceases registered around men, women;

C. Statistical models

The causality ratios among different demographic indices can be quantified and analysed with the help of correlation [7]. The obtained information offers the possibility of knowing the following aspects:

- The existence of the causality ratios among phenomena ;
- The contribution of each factor to the global variability of the effect phenomena;
- The intensity of the causes connections between the socio-economic phenomena and processes;
- The evolving tendencies of the correlation among phenomena.

In analyzing the correlation we take into account two main aspects:

- regression – which helps to determine the contribution of the determining factors to the variability of the effect phenomena by using and interpreting the regression coefficients of the different statistical-mathematical functions;

- the intensity of correlation – synthesised with the help of the correlation coefficients.

For the simple correlation, the first aspect can be pointed out with the help of the following functions: linear – for the causal-linear connections; higher parabolic, hyperbolic, exponential, logarithmic, semi-logarithmic, logistic, etc- for non-linear causal connections. The simple correlation can be linear and non-linear according to the form of the tendencies of causal connections.

The linear regression analysis uses the linear function with the following formula:

$$Y_x = a + bX \quad (10)$$

where
Y_x - empirical values;

a- values of the resulting variable Y determined outside the influence of the factorial variable X;

b- regression coefficient which synthesises the growth or the decrease of the Y variable corresponding to a growth or a decrease of the X variable which is equal to the unit;

X – values x1, x2, ... xn of the factorial variable X.

In practice, in order to determine the contribution of some of the influential factors (the factorial variable) for the variability of the dependent phenomenon, the **determination coefficient** is used with the following calculation relation [7]:

$$R^2 = \frac{[\operatorname{cov}(X,Y)]^2}{\sigma_x^2 \sigma_y^2} \quad (11)$$

where σ_x^2 and σ_y^2 represent the dispersion of the variable X, respectively Y, and $\operatorname{cov}(X,Y)$ represents a measure of the degree in which the variation of a variable suits with the variation of the other variable.

In practice R^2 is expressed in percentages.

The correlation coefficient *r* points out the intensity of the causal connection between the two variables. Both coefficients rely on the Pearson relation; the determination coefficient is the square of the correlation coefficient. Therefore:

$$r = \sqrt{R^2} = \frac{\text{cov}(X,Y)}{\sigma_x \sigma_y} \qquad (12)$$

where σ_x and σ_y represent the standard deviations for the X and Y variables.

The correlation coefficient r takes values between -1 and 1 with the following signification:

- values between 0 and 1 point out a direct correlation more and more intense as they are getting close to 1 ;

- values between 0 and -1 point out a reverse correlation more and more intense as they are getting close to -1;

- zero value points out the fact that there is no connection between these the two variables.

In order to test the signification of the correlation coefficient we are taking into account the verification of the null hypothesis (H_0) – which is carried out with the help of the test *Student (t)-* according to which there is no linear connection between the two variables. For this purpose, the parameter t_{calc} is calculated with the formula:

$$t_{calc} = r\sqrt{\frac{n-2}{1-r^2}} \qquad (13)$$

where

r – correlation coefficient;

n – number of value pairs *x* and *y*.

Value t_{calc} is compared with the tabular value from table *t (Student)*, for *n*-2 degrees of liberty and the determined signification degree (usually $\alpha = 0{,}05$).

The rejection region of the null hypothesis H_0, (for *n*-2 degrees of freedom) is:

$$t_{calc} > t_{\alpha/2,n-2} \text{ sau } t_{calc} < -t_{\alpha/2,n-2} \quad (14)$$

If H_0 is rejected, we can draw the conclusion with an assumed risk (usually 5%), that the value of the correlation coefficient does not equal 0, that is between the two variables there is a significant connection or in other words, the correlation coefficient is statistically significant.

If H_0 is accepted, he correlation coefficient is not significant and between the variables there is a casual connection.

III. STUDY ON THE POPULATION AGEING PHENOMENON

A. Analysis of the people pyramid and its implications

Starting from the socio-economic problems of the former communist countries [6] and after a close analysis of the population evolution, there have been pointed out the long term effects in these countries. Table 1 presents the evolution of the population in Romania according to age groups between 1990 and 2010. The population pyramid was made based on this data, Fig. 2.

TABLE 1. STRUCTURE OF ROMANIAN POPULATION ON AGE GROUPS

	Females		Males	
year	*1990*	*2010*	*1990*	*2010*
<5	886785	528013	924652	558232
5 to 9	840908	515289	879347	544950
10 to 14	965846	539363	1010941	566271
15 to 19	926797	617312	967764	646396
20 to 24	936256	845711	972155	880085
25 to 29	695718	792972	730443	830994
30 to 34	861549	861023	885855	908512
35 to 39	847369	821235	861305	850839
40 to 44	688201	832093	689118	844271
45 to 49	629281	624305	609852	613462
50 to 54	740820	776890	701320	728226
55 to 59	705496	749830	659262	671502
60 to 64	641048	598618	569872	509384
65 to 69	545966	509613	418864	396275
70 to 74	283694	547217	191709	382994
75 to 79	324529	429818	217727	279752
80 to 84	166702	272024	111738	162052
>85	73599	149767	48907	76896

Source: EUROSTAT

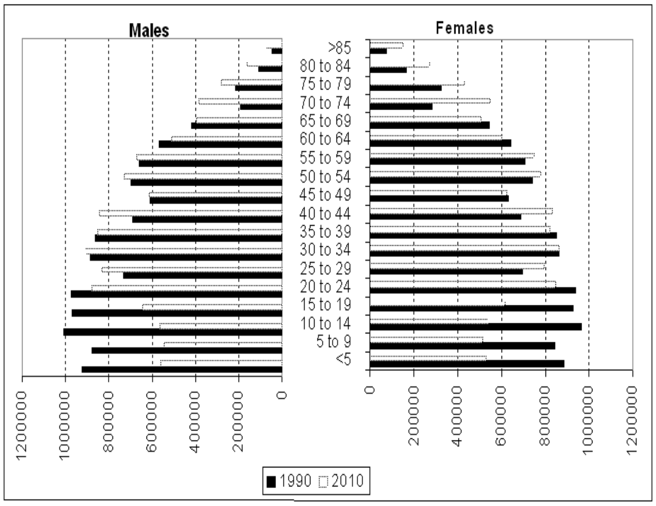


Figure 2. Population pyramid for Romania

A brief analysis of the type of pyramids for the two years shows that Romania has passed from a type 2 pyramid in 1990 when the ageing process when the ageing process was not too advances to a type 3 pyramid showing a population with advances symptoms of demographic ageing as a consequence of the significant decrease of fertility, announcing the process of de-population.

The ageing process is argues by the fact that the share of the population over 65 years old in the total number of population increase from 10,27% in 1990 to 14,94% in 2010. If in 1990 the percentage of 10,27% showed an ageing population, the one of 15% from 2010 overcomes substantially the limit of 12% (see previous section) which shows a demographic aged population.

At the same time, Fig. 2 shows that there was an ageing on the bottom of the pyramid due to the decrease of the birth rate. It is true that the share of the young population (0-14 years old) in the total number of population experienced a spectacular decrease from 23,73% in 1990 to 15,15% in 2010. The most obvious decrease can be noticed for the population from the age group 10-14 years old which is 44,06% smaller than that from 1990, followed by the <5 group (the age group younger than 5 years old) with a percentage of 40,03% and the 5-9 years old group with 38,36%. All these aspects show a decline of fertility in Romania in the 20 years from the fall of communism.

We notice that the principle of the people pyramid is fulfilled; according to this principle the adult population is relatively constant. There is a significant growth in the year 2010 for the 40-44 years old group (21,71%) which is explained due to the birth rate policy carried out by the government in the 70's.

On the whole, the female population of 65 years old and over in Romania has increased in 2010 by 513949 people as opposed to 1990, while the male population has increased by only 309024 people. Although the male ratio ((number of men/number of women)*100) shows a decreasing trend of 70,91 men to 100 women in 1990, of 68,01 to 100 women, in 2010, Romania holds one of the first places for this indicator. This fact is explained due to the life expectancy which is 77,1 years old for women and 69,6 years old for men 2009.

By taking over the information from the Eurostat database, the following results have been obtained for the other former communist countries which belong to EU [12].

The type of pyramid in Bulgaria is almost identical to the one in Romania, since it passed from a type 2 to a type 3. We can say that the ageing process is stronger in Bulgaria because the share of the aged population in the total number of population was of 12,96% in 1990 and of 17,53% in 2010, so in 1990 the population was already demographically aged. These high values explain why the top of the pyramid has a flattened form than the one of Romania.

In this case we also have an ageing on the bottom of the pyramid, the share of the young population (0-14 years old) in the total of population evolving from 20,54% in 1990 to 13,57% in 2010, that is a decrease of 6,97 percentage points. The largest decrease was registered for the age group 0-14 years old with a percentage of 51,24%.

The female population over 65 de years old, has increased in 2010 by 151057 people as opposed to 1990, while the male population by 38568 people. The male ration shows a decrease from 78,73 man to 100 women in 1990 to 68,51 in 2010, the larges as compared to other countries (almost 10 percentage points). In spite of these, together with Romania it holds one of the first places for the ration of the number of men to 100 women. In consequence, we can say that the demographic evolution from these two countries is very similar.

For Lithuania, the type of the people pyramid is similar to the one from Romania and Bulgaria. The share of the population over 65 de years old in the total number of population has increased from 10,81% in 1990 to 16,05 % in 2010, which is a similar evolution to that of Romania, from an ageing population to a demographically aged one.

The ageing took place on the bottom of the pyramid, the share of the young population (0-14 years old) in the total number of population evolving from 22,57% in 1990 to 15% in 2010, that is a decrease of 7,57 percentage points. Here the largest decrease was registered for the age group 5-9 years old with a percentage of 45,36%.

Although the people of working age are relatively constant one can notice high differences in certain age groups. Therefore, the population in the age group 40-55 years old is smaller in 1990 due to the effects of the Second World War, while the population in the interval 20-40 years old in 2010 is smaller due to the migration phenomenon.

The feminine population over 65 years old in Lithuania has increased in 2010 by 92118 as opposed to 1990 while the male population has increased by 42829 people. This fact explains why the pyramid for 2010 is slightly asymmetrical to the right. The male ration shows a decrease from 52,54 men for 100 women in 1990 to 50,96 in 2010.

And for Latvia the population pyramid shows a transition from a type 2 pyramid to a type 3, which stand for an ageing process. The share of the population over 65 years old in the total number of population has increased from 11,82% in 1990 to 17,36% in 2010, that is an evolution similar to the one from Romania from an ageing population to a demographically aged one.

The ageing took place on the bottom of the pyramid, the share of the young population (0-14 years old) in the total number of population evolved from 21,43% in 1990 to 13,75% in 2010, that is a decrease by 7,68% percentage points. Here the largest decrease was registered for the age group 10-14 years old (the same as for Romania and Bulgaria) with a percentage of 44,27%)

The female population over 65 years old from Latvia increased in 2010 by 44630 as opposed to 1990, more than the male population which registered a growth of 30189 people. The masculinity ratio shows a growth from 44,44 men for 100 women in 1990 to 48,37 in 2010.

The evolution of the ageing process in Estonia is much more similar to that from Romania and Bulgaria. The pyramid type is the same, the share of the population over 65 de years old in the total number of population has increased from 11,56% in 1990 to 17,07 % in 2010, that is an evolution from an ageing population to a demographically aged one.

The share of the young population (0-14 years old) in the total number of population evolved from 22,29% in 1990 to 15,14% in 2010, that is a decrease of 7,15 percentage points, meaning an ageing on the bottom of the pyramid. The largest decrease was registered for the age group 10-14 years old (the same as in Romania, Bulgaria and Latvia) with a percentage of 44,99%.

In this country the female population over 65 de years old has increased in 2010 by 26789 people as opposed to1990, more than the male population which registered a growth of 20359 people, the masculinity ratio showing a growth from 43,35 men to 100 women in 1990, to 49,05 in 2010.

Poland is one f the countries where the ageing process is on the bottom of the pyramid. The share of the population over 65 de years old in the total number of population has increased from 9,95% in 1990 to 13,52 % in 2010, which is a similar evolution to the one of Romania, from an ageing population to a demographically aged one. The growth in percentage points is of 3,57, which is an average value as compared to the other countries. The larges growth was registered by Latvia with 5,53 pp, and the smallest by Slovakia with 1,99 pp.

The share of the young population (0-14 years old) in the total number of population went down from 25,27% in 1990 to 15,15% in 2010, that is a decrease of 10,12 percentage points meaning the largest decrease in the analysed countries. The largest decrease was registered for the age group 5-9 years old (the same as in Lithuania) with a percentage of 47,25%.

The principle of people pyramid according to which the adult population is constant is not complied in Poland, this group of population increased by 10,48% in 2010 as opposed to 1990. One can notice high disparities in certain age groups. Therefore, the population from the age group 45-55 de years old from 1990 is smaller than that from 2010 due to the effects of the Second World War.

The female population over 65 de years old has increased in 2010 by 851586 people as opposed to 1990, more than the masculine population which registered a growth of 524221 people, the masculinity ratio showing a decrease from 60,03 men to 100 women in 1990, to 60,44 in 2010.

In Hungary we are also dealing with an ageing process. The share of the population over 65 de years old in the total number of population has increased from 13,24% (the highest level) in 1990 to 16,61 % in 2010, meaning that the population was already aged in 1990 and the phenomenon increased in 2010 with 3,37 pp.

The share of the young population (0-15 years old) in the total number of population went down from 20,54% in 1990 to 14,75% in 2010, that is a decrease of 5,79 percentage points meaning the slightest decrease in the analysed countries. The largest decrease was registered for the 5-9 years old age group (the same as in Lithuania and Poland) with a percentage of 26,50%.

The principle of people pyramid according to which the adult people of working age remains constant is complied in Hungary, the population being identical in the two analyzed years (the only country which does not registers a growth or a decrease).

The female population over 65 years old has increased in 2010 by 208690 people as opposed to 1990, more than the masculine population which registered a growth of only 80871 people, the masculinity ratio showing a decrease from 62,31 men to 100 women in 1990, to 57, 65 in 2010. An interesting aspect for this country is the fact that the number of people above 85 years old has doubled as opposed to 1990.

The Czech Republic also presents a demographic ageing process. The form of the pyramid evolves from a type 2 to a type 3. The share of the population over 65 de years old in the total number of population has increased from 12,47% in 1990 to 15,22% in 2010, meaning that the population was demographically aged in both analysed years. The growth in percentage points is of 2,75 one of the smallest growth , second after Slovakia.

The share of the young population (0-14 years old) in the total number of population went down from 21,74% in 1990 to 14,22% in 2010, that is a decrease of 7,52 percentage points, representing the smallest decrease as compared to other countries. The largest decrease was registered for the age group 10-14 years old with a percentage of 48,66%. It is remarkable the fact that in the last five years, there has been an increase in the birth rate, the bottoms of the two pyramids from the years 1990 and 2010 tend to unify. So for the age group <5 years old we have a decrease by 13,76%, the smallest as compared to the other countries..

The adult population has increased by 8,74%, which is explainable because the Czech Republic is one of the few countries which registered a growth in the number of population in 2010 as opposed to 1990. This is a unique situation as compared to the other analyses; the number of female population over 65 de years old is almost identical with the number of the male population (153206, respectively 153655 people). The masculinity ration shows a growth from 60,48 men to 100 women in 1990, to 66,85 in 2010. In this case, similarly to Hungary, the number of people over 85 years old almost doubled in 2010 as opposed to 1990, supporting one more time the demographic ageing phenomenon.

In Slovakia there has been an ageing on the bottom of the pyramid, meaning that the birth rate decreased. The share of the young population (0-14 years old) in the total number of population, experienced a tremendous decrease from 25,45% in 1990 to 15,32% in 2010 that is a decrease by 10,13 percentage points. This country together with Poland registered the largest decrease as compared to the other countries. The most obvious decrease is noticed for the population from the age group 5-9 years old, which is 42,25% smaller than the one from 1990, followed by the age group 10-14 years old with a percentage of 39,24% and the group <5 years old with 32,72%. Al these facts show a decline of fertility with the observation that in the period 2006-2010 there has been noticed an improvement of this indicator.

We notice that the principle of the people pyramid is not complied, according to which the people of working age (15-64 years old) remains relatively constant. The number of population has increased from 3398783 in 1990 to 3928471 in 2010, that is a percentage of 15,58%, the largest in the 10 analysed countries. The most spectacular growth was registered for the age groups 45-59 years old, and the smallest for the groups 15-44 years old, phenomenon which can be explained due to the decrease in the young population. Similar processes of growth were registered in Slovenia, Poland, Hungary and the Czech Republic.

The share of the population over 65 de years old in the total number of population has increased from 10,27% in 1990 to 12,26 % in 2010, that is an evolution similar to the one from Romania, from an ageing population to a demographically aged one. The growth in percentage points is of 1,99 that is the smallest value as compared to the other countries.

On the whole the female population over 65 years old from Slovakia has increased in 2010 by 91227 people as opposed to 1990, while the masculine population has increased by only 30992 people, meaning that the female population increased 3 times as compared to the masculine population. The masculinity ration shows a decrease from 44,75 men for 100 women in 1990 to 39,4 men in 2010.

An in Slovenia there has been an ageing process on the bottom of the pyramid due to the decrease in the birth rate. The share of the young population (0-14 years old) in the total number of population experienced a decrease from 20,95% in 1990 to 14,03% in 2010 that is a decrease of 6,92 percentage points. This country has registered one of the smallest decreases, being the second after Hungary. The most obvious decrease is noticed for the population in the age group10-14 years old, which is 38,01% smaller than the one from 1990, followed by the group 5-9 years old with a percentage of 35,76% and the group <5 years old with 18,15%. All these show a decline of the birth rate with the observation that in the period 2006-2010 this indicator has considerably improved.

The principle of people pyramid according to which the adult population is relatively constant, is fulfilled in this case. The number of population increased from 1366532 in 1990 to 1421436 in 2010, that is a percentage of 4,01%, Slovenia being one of the few countries which registered growth for this type of population.

The share of the population over 65 de years old in the total number of population increased from 10,60% in 1990 to 16,53 % in 2010, that is an evolution from an ageing population to a demographically aged one. The increase of 5,93 percentage points, that is the highest as compared to the other countries. These considerations show a strong ageing process.

On the whole the female population over 65 de years old from Slovenia has increased in 2010 by 67039 people as opposed to 1990, while the masculine population has increased by only 59620 people. The masculinity ratio shows a decrease from 34,37 men for 100 women in 1990, to 30,9 in 2010.

From the analysis we carried out above, we can distinguish the following common features regarding the demographic evolution of the 10 EU member countries in the period 1990-2010:

- The share of the young population (0-14 years old) in the total number of population has significantly increased;
- The share of the old population (65 de years old and over) in the total number of population has increased;
- The number of old women is larger than that of the men;
- The ageing process took place on the bottom of the pyramid.

In these general tendencies there are a series of disparities between the countries on the background of different socio-economic situations.

Therefore for the young population, age group <5 years old, the Czech Republic and Slovenia are registering the highest decreases (13,76% respectively 18,16%). These figures are correlated with the ones from the age groups 5-9 years old and 10-14 years old and show us that in these countries there is going to be a process of rejuvenation of the population. Taking the Czech Republic as an example we analysed the evolving trend for the age <5 with a parabolic model and we made a prediction for the next 6 years. The results, Fig. 3, show us the fact that around 2014 there will be a restoration of the number of people in the age group <5, from 1990.

On the opposite pole there is Latvia and Romania which announce a long term demographic ageing process due to the spectacular decreases (45,25% respectively 40,03%). For the rest of the countries we notice a shy beginning of population rejuvenation, but the process will last for a long time because the number of young people has significantly decreased.

Analysing the share of the old population in the total number of population we can notice that countries as Bulgaria, the Czech Republic and Hungary used to have in 1990 a population which was already demographically aged (shares higher than 12%). The other countries had an ageing population, with a special notice for Poland which registered the smallest share (9,95%) in that year. Expressed in percentage points, the largest growth were registered in Slovenia (5,93) and the Baltic Countries (approx. 5,5) and the smallest in the Czech Republic (2,75) and Slovakia (1,99).

In all these countries, the number of old female population has increased as compared to the number of men. For this chapter on the first place there is Hungary and Slovenia where the number of women is 3 times higher than the number of men, and on the opposite pole the Czech Republic where the two populations have equal values.

Comparing the level of decrease for the young population with the level of growth for the old population we can draw the conclusion that there has been an ageing process in all countries and it took place on the bottom of the pyramid. (see Def. 2).

B. The Analysis of the Mortality Rate

The mortality rate is an indicator which is used in the analysis of the ageing process of the population. Taking into account the subject of this article we have chosen the mortality rate on age and sex groups for the age group 65 de years old and over. The number of this population on the whole and sexes as well as the mortality rate is presented in Table 2 and Fig. 4.

The analysis of the data in Table 2, shows us the fact that in all countries the number of old population has increased in 2009 as opposed to 1990.

Fitted Trend Equation

Yt = 744734 - 40436,5*t + 1476,72*t**2

Forecasts

Period	Forecast
22	569865
23	595881
24	624850
25	656773
26	691649
27	729479

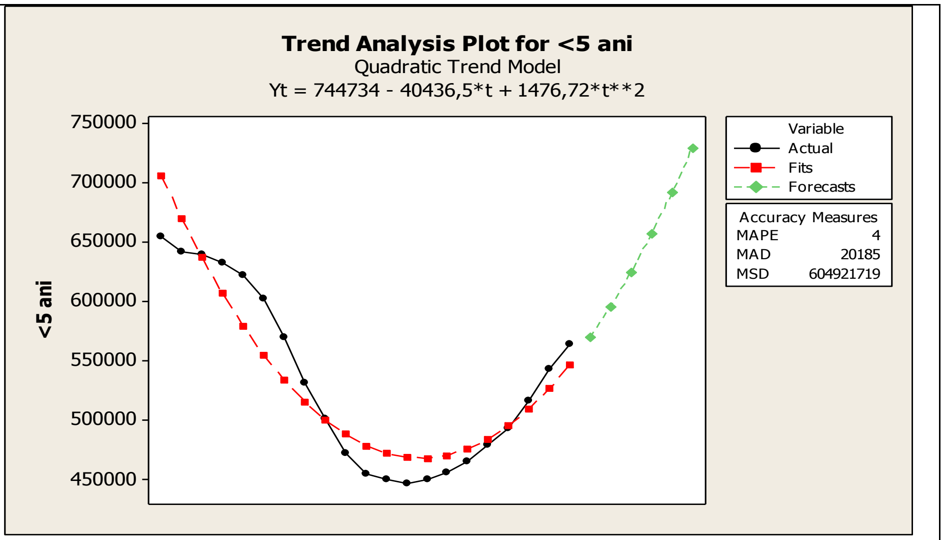


Figure 3. Evolving trend of the age group <5 for the Czech Republic

The largest percentage for the total number of people is registered by Slovenia with 57,85% followed by Poland with 35,94% and Romania with 34,19%. On the opposite pole there is Hungary with 19,39% and on the last place Bulgaria with only 16,51%.

The situation on sexes is the same as in the case of the total number of women but there are interesting changes on the bottom. Therefore, the smallest percentage (16,32%) goes to the Czech Republic while Bulgaria overcomes countries as Estonia, Latvia and Hungary. In the case of men, the first place goes to Slovenia with a high growth (77,39%) followed by Estonia and Poland. On The last places there is Hungary and Bulgaria. (13,61% respectively 7,89%).

TABLE 2. DATA OLD POPULATION

	Males				**Females**			
	No. population		***No. deceases***		***No. population***		***No. deceases***	
	1990	***2009***	***1990***	***2009***	***1990***	***2009***	***1990***	***2009***
BG	501	540	38	38	636	784	38	43
CZ	487	620	42	37	805	936	52	45
EE	55	76	5	5	127	154	8	7
LV	97	128	7	9	218	263	14	13
LT	138	182	11	13	262	354	14	17
HU	527	599	44	41	846	1041	53	52
PL	1420	1940	112	122	2365	3206	137	147
RO	989	1299	72	88	1394	1899	84	98
SI	74	131	5	6	138	203	7	8
SK	217	244	17	16	326	410	19	21

Source: EUROSTAT

With respect to the total number of old deceased people in the three countries (The Czech Republic, Hungary and Slovakia) there have been registered decreases in 2009 as opposed to 1990, the most significant being in the Czech Republic of 13,44%.The rest of the countries have registered growth on the first place being Romania with 22,69%.

The number of female deceased persons has decreased in the Czech Republic, Estonia, Latvia and Hungary, the highest decrease of 13,52% being in Estonia. The rest of the countries experienced growth, on the first place being Romania with 16,96%.

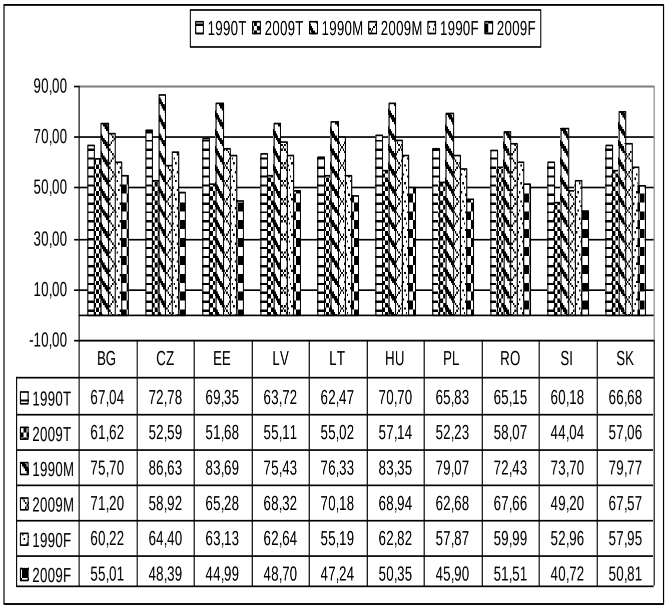
With respect to men there have been registered decreases in the Czech Republic (13,44%), followed by Hungary and Slovakia with 6,03% respectively 4,92%. The other countries experienced growth, Romania being still on the first place 22,69%.

The mortality rate is analysed according to the information from Fig. 4 and in the context of the life expectancy Fig. 5. The mortality rate on the entire old population has decreased in all 10 countries but with some particularities we are going to discuss. The highest values are in the Czech Republic, Estonia and Slovenia and on the last places Romania, Bulgaria and Lithuania. The particularity lies in the fact that the Czech Republic and Estonia have experienced decreases both in the number of deceased people (with 12,97% respectively 5,84%) and for the mortality rate (with 27,74% respectively 25,47%). In this context if we take into account the fact that the number of old population has increased in 2009 by 20,44% respectively 26,34% we can say that in there is a demographic ageing process in the two countries and a high life expectancy. For this indicator the Czech Republic has registered a growth from 71,26 in 1990 to 76,6 in 2009, that is a growth of 5,34 percentage points, being on the last place and Estonia with a life expectance of 74,5 years old in 2009 has a growth of 4,76 years.

Slovenia had a different evolution as compared to the two countries. The decrease of the mortality rate was higher (26,82%) but the number of deceased people has substantially raised by 15,52%. This negative status was cancelled by the spectacular growth of the old population by 57,85% in 2009 as opposed to 1990, fact which has led to the demographic ageing and to a life expectancy of 78,5 years old in 2009, the highest of the analysed countries.

The high values of the life expectancy in Slovenia and the Czech Republic are argued by the fact that the population from the age groups 80-84 de years old and over 85 de years old has increased in 2009 by almost 80% as opposed to 1990.

Romania and Bulgaria have one of the smallest life expectancies, of 73,3 or 73,4 years. We reached this result in a different way. Romania has experienced a slight decrease of the mortality rate (10,87%) but a high increase in the number of deceased people (19,60%) which was cancelled by a substantial growth of the number of old people (34,19%). Bulgaria experienced a decrease in the mortality rate of 8,09% an increased in the number of deceased people of only 7,08% (as opposed 19,60% in Romania) and an increase in the old population of only 16,51% in 2009. Although the evolutions of the indicators have been contradictory, the demographic ageing process is present but with a small life expectancy. Lithuania is on the last place in what the life expectancy is concerned (72,5 years old) with an evolution close to the one of Romania.



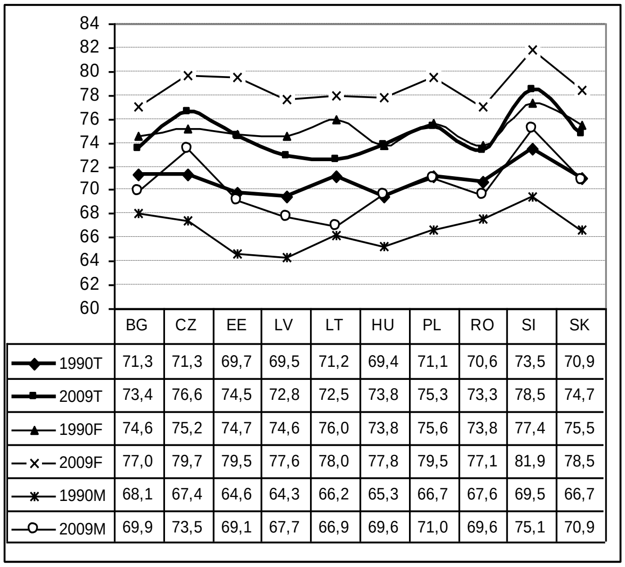
	BG	CZ	EE	LV	LT	HU	PL	RO	SI	SK
▤1990T	67,04	72,78	69,35	63,72	62,47	70,70	65,83	65,15	60,18	66,68
▩2009T	61,62	52,59	51,68	55,11	55,02	57,14	52,23	58,07	44,04	57,06
◨1990M	75,70	86,63	83,69	75,43	76,33	83,35	79,07	72,43	73,70	79,77
▨2009M	71,20	58,92	65,28	68,32	70,18	68,94	62,68	67,66	49,20	67,57
⊡1990F	60,22	64,40	63,13	62,64	55,19	62,82	57,87	59,99	52,96	57,95
■2009F	55,01	48,39	44,99	48,70	47,24	50,35	45,90	51,51	40,72	50,81

Figure 4. Mortality rate per total and per sexes for old population

The mortality rate on sexes for the old population has decreased in all 10 countries both for the female and the male population fact which shows a process of demographic ageing.

For the female population, the highest decreases for the mortality rate were registered in the Czech Republic and Estonia (24,85% and 28,74%), which correlated with the decrease in the number of deceased population (12,58% and 13,52%) and a moderate increase of the old population (16,32% and 21,36%) led to a process of demographic ageing for the female population and to a life expectancy of 79,7 respectively 79,5 years old, being overcome by Slovenia with 81,9 years old. As for the total of old population, Slovenia registered a high mortality (23,10%), but as compared to the other two countries an increase in the number of deceased persons (cu 13,36%) compensated by a high increase in the feminine population (47,42%). We have to notice for this type of population the life expectancy in Poland (79,5 years old) is identical with that from Estonia but the evolution of the indicators is not similar but is follows the trend of the one from Slovenia.

Romania and Bulgaria are on the last two places with a decrease of the mortality rate of only 14,3% respectively 8,65%. The other indicators have similar evolution with the one from the old population and led to a life expectancy on only 77,1 respectively 77,0 years old.



	BG	CZ	EE	LV	LT	HU	PL	RO	SI	SK
◆ 1990T	71,3	71,3	69,7	69,5	71,2	69,4	71,1	70,6	73,5	70,9
■ 2009T	73,4	76,6	74,5	72,8	72,5	73,8	75,3	73,3	78,5	74,7
▲ 1990F	74,6	75,2	74,7	74,6	76,0	73,8	75,6	73,8	77,4	75,5
× 2009F	77,0	79,7	79,5	77,6	78,0	77,8	79,5	77,1	81,9	78,5
✱ 1990M	68,1	67,4	64,6	64,3	66,2	65,3	66,7	67,6	69,5	66,7
○ 2009M	69,9	73,5	69,1	67,7	66,9	69,6	71,0	69,6	75,1	70,9

Figure 5. Life expectancy at birth

For the male population the highest decrease in the mortality rate was registered in Slovenia (with 33,25%) but also an increase in the number of deceases of 18,41%. This data correlated with the high increase in the number of old population (77,39%) make Slovenia the country with the highest life expectancy, 75,1 years old. On the second place there is the Czech Republic with 73,5 years old, but as compared to Slovenia there is also a decrease in the number of deceased persons. The evolution of the two countries is similar with the ones registered in the feminine population and old population.

Romania and Bulgaria are not on the last places, these being occupied by Estonia, Latvia and Lithuania. The surprising emergence of Estonia in this category is due to the fact that there has been a growth by 7,50% in the number of deceased people. Lithuania is on the last place for the life expectancy with 66,9 years old, under the circumstances of a mortality rate with a modest decrease of 8,06% but a strong growth of the number of deceased people (21,50%).

C. Demographic dependencies. Implications

The analysis of the demographic ageing process requires the study of the demographic dependency indicator [3, 8, 9], due to its social and economic implications.

Therefore the growth of the percentage of population over 65 years old and on the total determines the decrease of the share in the other age groups, and tends to create social and economic pressures determined by the change of the way of granting the resources in the society. In Table 3 we present the old age dependency rates [relation 5] for the years 1990 and 2010.

A high demographic dependency rate indicates the fact that a high number of beneficiaries of public health and retirement systems will be ”supported” by a low number of tax payers, the analysis of these aspect were done in [4]. Therefore the old people of working age will be ”burdened” to pay higher taxes and contributions which should provide the retired people a stable and sufficient income. In other words there will be a decrease in the living standard of the population.

TABLE 3. DEPENDENCY RATE FOR OLD PEOPLE

	People of working age (15-64)		**Old population (65+)**		**Rdv**	
	1990	***2010***	***1990***	***2010***	***1990***	***2010***
BG	5830075	5211619	1136266	1325891	19,49	25,44
CZ	6817371	7413560	1292022	1598883	18,95	21,57
EE	1038860	908466	181605	228753	17,48	25,18
LV	1780927	1549011	315390	390209	17,71	25,19
LT	2460639	2295339	399454	534401	16,23	23,28
HU	6870352	6873985	1373922	1663483	20,00	24,20
PL	24639820	27223082	3785663	5161470	15,36	18,96
RO	15319481	15003660	2383435	3206408	15,56	21,37
SI	1366532	1421436	211606	338265	15,48	23,80
SK	3398783	3928471	542915	665134	15,97	16,93

Source : EUROSTAT

The comparative analysis of the data from Table 2 shows a series of interesting aspects. Slovenia presents the highest growth for the dependency rate (8,31 percentage points) although the adult population has increased. It is the country with the highest pressure exerted by the old population on the people of working age. Here we can see a strong process of demographic ageing, fact which was underlined above, where we showed that the share of the old population increased in 2010 by 5,93 percentage points as opposed to 1990 (the largest growth as compared to the analysed countries).

For the Czech Republic, Poland and Slovakia there is also a growth in the number of people of working age, but the dependency rates have increased with less percentage points(2,62 for the Czech Republic, 3,60 for Poland and 0,96 for Slovenia), which shows us a moderate process of demographic ageing. Hungary is also in this category – with an adult population almost identical it registered a growth in the dependency rate by 4,20 percentage points.

Romania and Bulgaria show a decrease in the number of people of working age and at the same time the dependency rates have significantly increase in percentage points by 5,81 and 5,95. We can conclude that in these countries there is a high pressure of the old population on adult population. There is a demographic ageing process in these countries as well.

The most difficult situation is registered for the Baltic Countries which register decreases for the number of people of working age and high growth for the demographic dependency rate, fact which indicates a strong ageing process of the population and a strong pressure on the adult population. The causality ratios between the people of people of working age (PVM) and the population over 65 years old (P65+) can be quantified and analysed with the help of regression and correlation.

The data to be analyzed, Fig. 6, are taken from EUROSTAT and represent the 1990-2010. Using the linear regression where PVM is the factorial variable and P65+ the resulting variable and the analysis and data processing program MINITAB 14.1, we will obtain the following information for Romania:

Regression Analysis: P65+ versus PVM

```
The regression equation is

P65+ = 20486378 – 1,15 ·PVM
Predictor    Coef        SE Coef       T        P
Constant     20486378   4860702       4,21    0,000469
PVM          -1,15       0,320301     -3,61    0,001834
R-Sq = 40,7%
```

Correlations: P65+; PVM
Pearson correlation (r)= - 0,63

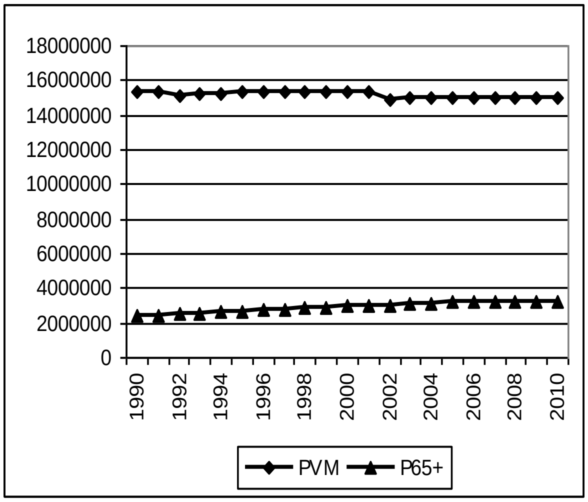


Figure 6. Evolution of indicators PVM and P65+ in Romania

The value of the determination coefficient (R-Sq), points out the fact that PVM (number of people of working age) influences P65+ (number of old population) in proportion of 40,7%.

At the same time the value of the correlation coefficient *r* (-0,63) shows a reverse correlation with average intensity. The signification of this result is noticed with the help of the test *Student (t)*, where t_{calc} calculated with relation 13 has the value -3,617. The tabular value of *t* for 21-2 degrees of freedom and α/2=0,025 is: 2,093 (respectively -2,093).

As the value of t_{calc} is outside the interval of the tabular values it means that *r* is significant, that is there is a significant connection between the two variables or in other words the correlation coefficient is statistically significant.

We can notice that the value of the constant parameter *a* is very high (20486378) and that the determination coefficient has a modest value (0,407) meaning that other factors are influencing the number of old population. For example we might use the multiple regression with two factorial parameters (PVM and the life expectancy-SV) obtaining the following results:

```
The regression equation is
P65+ = - 3534847 - 0,285 PVM + 150851 SV

   Predictor     Coef          SE Coef      T        P
   Constant  -3534847      7181243    -0,49    0,629
   PVM       -0,2846       0,3303     -0,86     0,400
   SV           150851      38731        3,89     0,001

R-Sq = 67,9%

Correlations: PVM; SV; P65+

            PVM     SV
   SV     -0,679
   P65+  -0,639     0,816
```

In this case the determination coefficient has substantially increased showing the fact that PVM and SV are influencing P65+ in a proportion of 67,9%. The partial correlation coefficient for the life expectancy ($r_{P65+,SV}$) has a very high positive value (0,816), meaning that between the variations of the two variables (P65+ and SV) there is a direct strong connection. The partial correlation coefficient for PVM ($r_{P65+,PVM}$) has a moderate negative value (-0,639), meaning that between the variations of the two variables (P65+ şi PVM) there is a reverse connection of average intensity. The compound correlation coefficient $r_{SV,PVM}$, for the two factorial variables-PVM şi SV- has a value of -0,679 pointing out a reverse moderate intensity connection.

Working in a similar way for the other countries we obtain a series of information regarding the correlation between the two variables, Table 4.

TABLE 4. PARAMETERS OF THE REGRESSION EQUATION AND THE CORRELATION COEFFICIENTS

	Regression equation	**R-Sq (%)**	**r**	t_{calc}
BG	P65+ = 2873139 -0,28 ·PVM	70,6	-0,84	-6,757
CZ	P65+ = -1569689+0,41·PVM	92,7	0,96	15,625
EE	P65+ = 491010,1 -0,30·PVM	65,6	-0,81	-6,032
LV	P65+ = 890211,1 -0,32·PVM	79,0	-0,88	-8,462
LT	P65+ =2205327 -0,73·PVM	87,8	-0,88	-11,718
HU	P65+ =5652582 -0,59·PVM	6,45	-0,25	-1,144
PL	P65+ =-9843770 +0,55·PVM	99,7	0,99	44,991
RO	P65+ = 20486378-1,15·PVM	40,7	-0,63	-3,617
SI	P65+ = -3274244+2,54·PVM	94,7	0,97	18,577
SK	P65+ = -130823 +0,19·PVM	96,9	0,98	24,56

From the analysis of the data from Table 4 from the point of view of the determination coefficient (R-Sq) we can notice that Hungary has the smallest coefficient, adult population has an insignificant influence on the number of old population, which means that other factors have influenced this number. For example the growth of the living standard which led to the growth of the life expectancy, Hungary registering an important performance for this chapter. The life expectancy for men has increase from 64,29 years old in 1990 to 68,7 years old in 2009, and for women from 72,84 years old to 76,8 years old; with an average of 4,2 years it holds the third place after the Czech Republic and Slovenia.

Romania is on the second place in this classification with a determination coefficient of only 0,407 which shows that the influence of other factors on the number of old people is pretty high (59,3%). This situation is pointed out by the very high value (20486378) of the parameter *a* (intercept) for the regression equation.

On the opposite pole there is a group of countries (The Czech Republic, Poland, Slovenia and Slovakia) where the influence of the active population on the old population is above 90%, which is other factors have a slight influence. Poland is remarkable with an influence percentage of 99,7% and shows us that the number of old population was determined exclusively by the people of working age number.

There is a group of countries between these extremes (Bulgaria, Estonia, Latvia, Lithuania) where the determination coefficient has a moderate value, suggesting the fact that other factors have a pretty high influence on the number of old population.

The intensity and sense of the correlation between the two variables is analysed through the correlation coefficient. The intensity level was refined on intervals of values from a strong, functional connection to a weak one or to the inexistence of a connection.

Bulgaria, Estonia, Lithuania, Latvia, Hungary and Romania present a reverse correlation between the variables with the following intensities: Hungary, very weak intensity (-0,25), almost non-existing, Romania has an average intensity and the other countries a strong intensity.

The Czech Republic, Slovakia and Slovenia present a direct correlation with high values of the correlation coefficient (above 0,95) indicating a functional connection between the two variables.

The significance of the above results is pointed out with the help of the statistical test *Student(t)*. Therefore, all analysed countries except for Hungary have the value of t_{calc} outside the interval of the tabular values that is the null hypothesis is rejected and between the variables there is a causal connection. Hungary has a value of t_{calc} of -1,144 situated in the interval of the tabular values, meaning that this null hypothesis is accepted, in other words there is no causality relation between the two variables.

IV. Conclusions

This study pointed out the way in which different population parameters have evolved in the former-communist countries and their implications of the social environment. It was notices that some of the data was influenced by the economic and political evolution and in some cases there were major interferences between the evolution of the population and that of the involved parameters. The obtained results show the fact that in the former-communist countries we see a growing ageing phenomenon which exists in the other EU countries, but the determining factor for the first group of countries is the severe decrease of the birth rate after the removal of the restrictive policies regarding abortions (after 1990).All these data rely on official EU sources (EUROSTAT) and their analysis has been made using mathematical analysis methods for demographic statistics and analysis and statistical data processing software. Without the claim of having discussed all the interpretative valences of the phenomenon, our intention is to further study the implications on the economies of the states with which of studied countries have economic relations and their social motivations, using non-linear models from the statistical mathematics.

REFERENCES

[1] R. H. Binstock, K. G. Linda (Editors), "Handbook of Aging and the Social Sciences", 7th Edition, Academic Press, 2010

[2] G. Calot, J.-P. Sardon, "Methodology for the calculation of Eurostat's demographic indicators", Eurostat, 2003

[3] H. Engelhardt, H.-P. Kohler, A. Prskawetz, "Causal Analysis in Population Studies: Concepts, Methods, Applications", Springer, 2009

[4] J. Grant et all, "Low Fertility and Population Ageing Causes, Consequences and Policy Options", RAND Europe Report of European Commission, 2004

[5] A. Haupt, T. Kane, "Population Handbook", Population Reference Bureau, 5rd edition, Washington, 2004

[6] M. M. Howard, "The Weakness of Postcommunist Civil Society", Cambridge University Press, 2002

[7] G. Keller, B. Warrack, "Statistics for Management and Economics", Academic I. Pub Inc, 2006

[8] W. Lutz, "Recent Demographic Trends in Europe and the World", European Demographic Forum, 2008

[9] R. Moffitt, "Remarks on the Analysis of Causal Relationships in Population Research, Demography", no. 42, 2005

[10] L. A. Morgan, S. R. Kunkel, "Aging, Society, and the Life Course", Fourth Edition Springer Publishing Company, 2011

[11] S. H. Preston, P. Heuveline, M. Guillot, "Demography: Measuring and Modelling Population Processes", Blackwell Publishing. 2001

[12] ***http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database

AUTHORS PROFILE

Dorel Săvulea: He is Lecturer at University of Craiova, Faculty of Exact Sciences, Department of Informatics. His area of expertise is in the domain of economic process analysis.

Nicolae Constantinescu: He is Associate Professor at University of Craiova, Faculty of Exact Sciences, Department of Informatics and his area of expertise is mathematically models of cryptography and statistics analyses.

Pattern Recognition-Based Environment Identification for Robust Wireless Devices Positioning

Nesreen I. Ziedan
Computer and Systems Engineering Department
Faculty of Engineering, Zagazig University
Zagazig, Egypt

***Abstract*—There has been a continuous increase in the demands for Global Navigation Satellite System (GNSS) receivers in a wide range of applications. More and more wireless and mobile devices are equipped with built-in GNSS receivers; their users' mobility behavior can result in challenging signal conditions that have detrimental effects on the receivers' tracking and positioning accuracy. A major error source is the multipath signals, which are signals that are reflected off different surfaces and propagated to the receiver's antenna via different paths. Analysis of the received multipath signals indicated that their characteristics depend on the surrounding environment. This paper introduces a machine-learning pattern recognition algorithm that utilizes the aforementioned dependency to classify the multipath signals' characteristics and identify the surrounding environment. The identified environment is utilized in a novel adaptive tracking technique that enables a GNSS receiver to change its tracking strategy to best suit the current signal condition. This will lead to a robust positioning under challenging signal conditions. The algorithm is verified using real and simulated Global Positioning System (GPS) signals with accurate multipath models.**

Keywords-component; GPS; GNSS; machine learning; pattern recognition; PCA; PNN; multipath.

I. INTRODUCTION

A Global Navigation Satellite System (GNSS) [1, 2] is a radio navigation system that employs spread spectrum techniques to transmit ranging signals and navigation data. The ranging signals are used by a GNSS receiver to identify the visible GNSS satellites and measure the distance between the visible GNSS satellites and the GNSS receiver. The measured distances are used with the navigation data to solve the navigation equation to determine the user's 3-diemntional position and velocity. Examples of GNSS systems are the US Global Positioning System (GPS), the Russian GLONASS, and the European Galileo Navigation System.

GNSS receivers perform three main functions: signal acquisition, signal tracking, and navigation message decoding. Signal acquisition identifies the visible satellites and provides rough estimates of the Doppler frequency, f_d, and the ranging code delay, τ. Signal tracking applies closed loop tracking techniques to provide continuous accurate estimates of the carrier phase, the Doppler frequency, the Doppler rate, and the code delay. Those estimates are used to measure the distance between the GNSS satellite and the receiver.

GNSS receivers can give positioning accuracy up to a few millimeters when the receiver is stable and has a clear view of the sky, where the Line-of-Sight (LOS) signal is received with strong power. However, in environments like urban, suburban, and indoor, the received signals suffer from attenuation and multipath errors because of the surrounding objects [3]. In addition, the user's mobility behavior can subject the receiver to changing and unstable signals dynamics. This leads to deterioration in the tracking and positioning accuracy.

Multipath signals are a major error source. They appear when the GNSS satellite signals are reflected off different surfaces and propagated to the receiver's antenna via different paths. This leads to the reception of several versions of the same signal, which causes tracking errors. Analysis of the received signals indicated that their characteristics depend on the surrounding environment [4, 5, 6, 7, 8]. Urban, sub-urban and indoor environments generate different characteristics, which include multipath signals' parameters like the number and duration of echoes and signals power, and LOS signal's parameters like amplitude fluctuation, Doppler shift and rate. Different tracking strategies are needed for each environment to mitigate multipath errors and maximize the tracking performance.

There have been numerous tracking strategies that are optimized for specific signal condition or environment. For example, conventional tracking techniques [1, 2] are used with strong signals. Kalman Filter based techniques are used with weak signals [3, 9]. Tightly-coupled GNSS with Inertial Navigation System (INS) techniques are used with weak interrupted signals or blocked signals [10]. Open-loop batch processing, and combined batch and sequential processing techniques are used in high dynamic applications [11]. Particle Filter-based techniques are used for tracking in multipath environments [12, 13, 14].

A GNSS receiver is usually tuned to one tracking technique, and there have been no methods that enable a receiver to change its tracking strategy based on the surrounding environment. This paper introduces a machine-learning pattern recognition algorithm to identify the surrounding environment, and hence enable the implementation of a tacking strategy selector that adaptively changes the tracking strategy to best suit the current signal condition.

The LOS and multipath signals' patterns of each possible environment are represented by a class. The introduced algorithm is structured into several channels, each of which is tuned to one of the classes. The channels are trained on sets of patterns from each class, and then they are used to classify new unclassified patterns. The proposed pattern recognition algorithm performs two main functions, which are feature extraction and pattern classification. Feature extraction is the process of learning the distinctive characteristics of the data and removing redundant data. This is done to get a compact and robust representation of the distinctive features of each class, thus reducing the processing overhead and memory requirements without degrading the classification performance. Feature extraction, which is used in image and face recognition [15, 16, 17, 18, 19, 20, 21, 22, 23], is performed using a Principal Component Analysis (PCA) approach. Pattern classification is the process of building neurons that capture the dominant features shared by different realizations of each class, and then classifying new patterns into one of the classes. Neurons are computational units that are connected by weighted links. Pattern classification has many applications, like radar detection and remote sensing [24, 25, 26]; it is performed here using a multi-layer feedforward Probabilistic Neural Network (PNN) approach [27, 28, 29, 30, 31,32].

The proposed tracking strategy selector module utilizes both closed loop tracking techniques and open loop tracking techniques to accommodate various patterns. Open loop tracking techniques are activated in unstable or rapidly changing signal conditions, while closed loop tracking techniques are activated in relatively stable or light multipath environments. In addition, based on the environment classification, the activated technique adjusts its parameters to achieve reliable tracking performance.

The rest of this paper is organized as follows. Section II presents some signals patterns that appear in urban and suburban environments. Section III presents the proposed pattern recognition algorithm. Section IV discusses the adaptive tracking concept. Section V presents some testing and results to verify the algorithm performance. Section VI concludes the paper.

II.LOS And Multipath Patterns

The received GNSS signal consists of the LOS signal and N_{MP} multipath signals. The down-converted sampled received GPS C/A signal can be expressed as

$$\begin{aligned} r &= A\, d_\tau\, C_\tau \cos\left(\theta_n + 2\pi (f_{IF} + f_d) t + \pi \alpha t^2\right) \\ &+ A \sum_{m=1}^{N_{MP}} \Psi_m\, d_{\tau_m} C_{\tau_m} \\ &\quad \cos\left(\theta_n + 2\pi (f_{IF} + f_d) t + \pi \alpha t^2 + \phi_m\right) + n \end{aligned} \tag{1}$$

Where, A is the signal amplitude. d is the navigation data. C is the ranging code. f_{IF} is the intermediate frequency (IF). f_d is the Doppler shift. α is the Doppler rate. θ_n is the phase. τ is the code delay. n is the measurement noise. Ψ_m is the attenuation in the multipath signal amplitude relative to the LOS signal amplitude. τ_m is the multipath signal delay. Φ_m is the multipath phase advance relative to the LOS signal.

The received signal is processed by the receiver to generate the integrated signal, which has the form

$$\begin{aligned} y_u(\Delta) &= \Lambda \gamma_{o_u} d_u R(\tau_{e_u} + \Delta) \operatorname{sinc}(f_{e_u} T_i)\, e^{j\theta_{e_u}} \\ &+ \Lambda d_u \sum_{m=1}^{N_{MP}} \gamma_{m_u} R(\tau_{e_u} + \Delta - \tau_{m_u}) \\ &\quad \operatorname{sinc}(f_{e_u} T_i)\, e^{j(\theta_{e_u} + \phi_{m_u})} + n_{\Delta_u} \end{aligned} \tag{2}$$

Where, Λ is a reference amplitude that is used to express any amplitude relative to it, e.g. the LOS amplitude is $A = \gamma_0 \Lambda$. $\gamma_m = \gamma_0 \Psi_m$. R(.) is the auto-correlation function. τ_{eu} is the code delay estimation error at time instance u. f_{eu} is the Doppler shift estimation error. θ_{eu} is the phase estimation error. Δ is a code delay relative to the estimated code delay of the LOS signal. T_i is the integration time.

The attenuation of the LOS signal and the number and distribution of the multipath signals depend on the surrounding environment. Elaborate studies exploring the effects of urban and suburban environments on real signals were presented in [4, 5, 6, 7, 8]. Identifying the distribution of the LOS and multipath signals will enable adjusting the tracking strategy or the tracking parameters to obtain the best attainable tracking performance under various signal conditions. The software provided in [33] is used to generate received signal patterns that typically appear in urban and suburban environments. Some of these patterns are shown in Figs. 1-3. Each figure shows two plots: a 3-dimensional (3-D) plot that expresses the pattern in time, delay, and power; and a 2-D plot that is a rotated version of the 3-D plot.

Fig. 1 shows a pattern generated in a suburban environment when a user is walking at a speed of 4 miles/hour. The pattern exhibits a strong LOS signal with light multipath signals that have weak power compared to the LOS signal. Fig. 2 shows a pattern generated in a suburban environment when a car is moving at a speed of 20 miles/hour. The LOS signal here is weaker than the LOS signal shown in Fig. 1, and the multipath signals are not much weaker than the LOS signal. The pattern exhibits changing characteristics over 5 seconds, and it can be divided into several sub-patterns. For example, from 0-2 seconds the LOS signal is stronger than the multipath signals, around 2 seconds the LOS signal appears to be completely blocked, from 2-3.5 seconds the LOS signal is weaker than the multipath signals, and around 3.5 seconds there are no multipath signals. Fig. 3 shows a pattern generated in an urban environment when a car is moving at a speed of 20 miles/hour. This pattern also exhibits changing characteristics. The LOS signal either suffers from complete blockage, or frequent interruption over a short period of time. The multipath signals in this urban environment are denser than the suburban cases.

The distribution of the surrounding objects and their shape and height directly contribute to the multipath signals (echoes) distribution. Indoor patterns are usually characterized by weak or blocked LOS signals. Obviously, different patterns will need different tracking strategies to achieve optimized tracking performance, and hence optimized positioning accuracy.

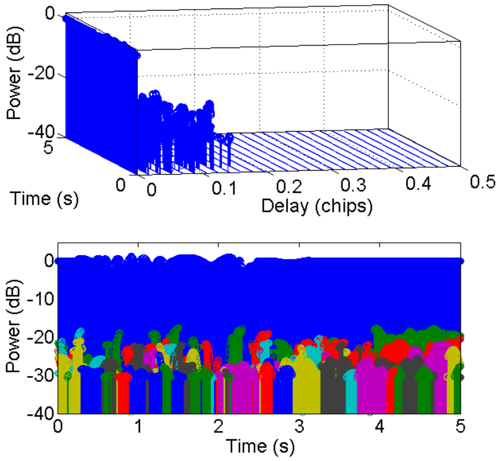


Figure 1. Multipath pattern in a pedestrian suburban environment.

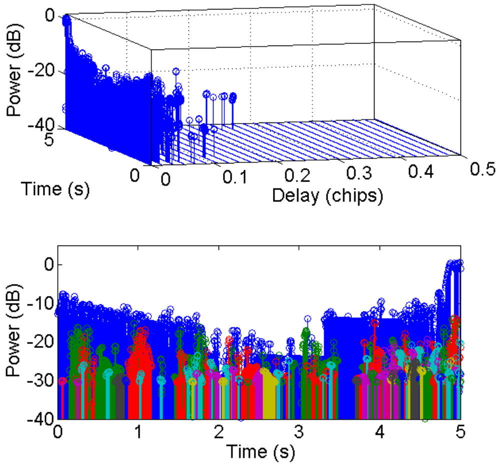


Figure 2. Multipath pattern in a car suburban environment.

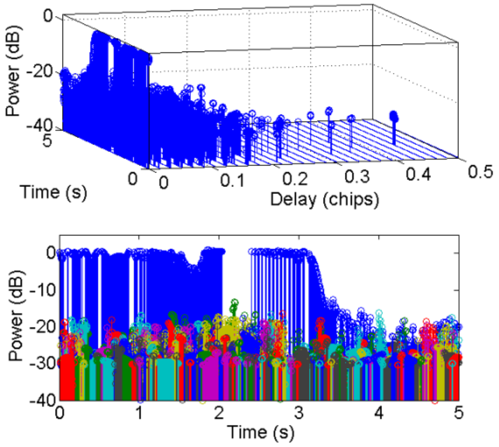


Figure 3. Multipath pattern in a car urban environment.

III. Multipath Pattern Recognition (MPR)

3-D patterns can be constructed from different parameters, like amplitude and phase. Each 3-D pattern is represented by a 2-D matrix, where each entry contains the value of a parameter at a code delay and a time instance. For example, a matrix that expresses the amplitude pattern over N_p code delays and N_{PNN} time instances is

$$\Omega_\gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1N_{PNN}} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2N_{PNN}} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{N_p1} & \gamma_{N_p2} & \cdots & \gamma_{N_pN_{PNN}} \end{bmatrix}$$

A PNN maps an input feature vector to an output class. The feature vector describes the similarities and differences between the features of the pattern in question and the dominant features of the training patterns. The process of learning the dominant features of the training patterns, and removing redundant data, is feature extraction.

Feature extraction is performed using PCA. In PCA, an eigen-decomposition is performed on the covariance matrix of the training patterns. The eigenvalues are sorted in a decreasing order, and then the eigenvectors that are associated with the largest eigenvalues, i.e. the dominant eigenvectors, are retained. A reduced-size pattern matrix is reconstructed with the dominant eigenvectors. The training patterns are divided into N_{cl} classes, where the characteristics of each class are selected based on the desired identification criteria. The criteria can be chosen to characterize the reflected signals (e.g. dense reflected signals, occasionally blocked LOS signal), a specific surrounding objects (e.g. trees, buildings), or a specific environment (e.g. urban, suburban). The classes should have distinctive characteristics relative to each other. The criteria and the number of steps, N_{PNN}, can be adjusted and optimized to fit the classification objective. The following will use the pattern matrix Ω_γ to explain the proposed algorithm.

Assume that there are a total of N_{tr} training patterns from all the N_{cl} classes. The PCA decomposition on the training matrices Ω_γ^k where k = 1, …, N_{tr}, is done as follows:

Each Ω_γ^k matrix is converted into one vector, by reading the matrix column by column.

The N_{tr} vectors are joined in one matrix, where each row contains one vector. So, a matrix of size N_{tr} (N_p N_{PNN}) is obtained; define it as Ω_γ.

The mean pattern is calculated. This is the mean of each column of Ω_γ. A row vector, Ω_γ^m, of size (N_p N_{PNN}) is obtained.

The mean pattern is subtracted from each row in Ω_γ to get a matrix defined as ξ_γ.

Eigen patterns (or eigen images) are patterns that characterize the similarities and differences between the training patterns. The eigen patterns are obtained using the training set, ξ_γ, and they are used in the classification process of unclassified patterns. Eigen patterns are the eigenvectors of the following covariance matrix: $\hat{\sigma}_{tr} = \xi_\gamma^H \xi_\gamma$. Where, $\hat{\sigma}_{tr}$ has a size of (N_p N_{PNN}) . (N_p N_{PNN}), and it will produce (N_p N_{PNN}) eigenvectors. It can be an intractable task to decompose a matrix of large size, so an alternative approach is used for decomposition. This is possible because the algorithm only needs a maximum of N_{tr} eigenvectors (those with the largest eigenvalues), and not all the (N_p N_{PNN}) eigenvectors. Getting the eigen patterns is done as follows:

- An alternative covariance matrix of size N_{tr} . N_{tr} is found as $\sigma_{tr} = \xi_\gamma\ \xi_\gamma^{\ H}$.
- The eigenvectors and eigenvalues of the covariance matrix σ_{tr} are calculated. The eigenvectors are a set of orthonormal vectors.
- The eigenvectors are sorted based on the associated eigenvalues. The N_{Etr} dominant eigenvectors are chosen, where $N_{Etr} \leq N_{tr}$. The chosen eigenvectors are appended to one matrix, defined as Γ_{tr}, which will have a size of N_{tr} N_{Etr}.
- The eigen patterns are obtain as $\lambda_{tr} = \xi_\gamma^{\ H}\ \Gamma_{tr}$, where, λ_{tr} has a size of (N_p N_{PNN}) . N_{Etr}. Those eigen patterns are orthonormal vectors, and they span N_{Etr}-dimensional subspace, instead of the original (N_p N_{PNN}) space.

The contribution of each eigen pattern in representing each training pattern is calculated by projecting the training patterns into the pattern space. This results in the following PCA weights: $W_{tr} = \lambda_{tr}^{\ H}\ \xi_\gamma^{\ H}$. Where, W_{tr} has a size of N_{Etr} . N_{tr}. Each column represents the projection of one training pattern into the pattern space, and each entry in a column represents the contribution of each eigen pattern in expressing the associated training pattern.

This process accomplishes two goals: expressing the patterns in terms of the dominant eigenvectors, i.e. the features that distinctly characterize the patterns; and, reducing the size of the training patterns.

The concept of classification of a new unclassified pattern is to find the class that best describes that pattern. The classification is based on finding the class with the minimum Euclidean distance to the projection of the new pattern. To classify a new pattern, $\Omega_\gamma^{\ n}$, it is first projected into the training pattern space as follows:

1) $\Omega_\gamma^{\ n}$ is converted into one vector, by reading the matrix column by column. A vector $\tilde{\Omega}_\gamma^{\ n}$ is obtained.

2) The mean pattern, $\Omega_\gamma^{\ m}$, is subtracted from $\tilde{\Omega}_\gamma^{\ n}$ to get a vector $\widetilde{\Omega}_\gamma^{n}$.

3) $\widetilde{\Omega}_\gamma^{n}$ is projected into the training pattern space (i.e. the PCA weights are calculated) as $W_p = \lambda_{tr}^{\ H}\ (\widetilde{\Omega}_\gamma^{n})^H$. Where. W_p has a size of N_{Etr} . N_{tr}.

4) The feature vector of the new pattern can be obtained as $\widetilde{\Omega}_\gamma^{p} = \lambda_{tr}\ W_p$.

The classification is done using PNN. The introduced PNN is a fully connected network that consists of four layers: input layer, pattern layer, summation layer, and output layer. Fig. 4 illustrates the PNN layers and their connections. The input layer has a number of units equal to the number of variables used in the classification process.

The training patterns are stored in the pattern layer, and they are used to construct the probability density function of each class. The summation layer employs a Bayesian approach to calculate the probability that a new unclassified pattern belongs to each class.

The output layer uses a competitive transfer function to choose the maximum probability and generate an output that indicates the class that the new pattern belongs to it. The introduced MPR algorithm is designed with a flexibility to learn new patterns, and adapt the performance based on new training data. Adding new patterns requires recalculating the weights W_{tr}.

The number of training patterns available for each class is defined as N_c, where c = 1,…, N_{cl}. The eigen patterns for each class are defined as $\lambda_{tri}^{\ c}$, where c is the class index, and i = 1, …, N_c. The PDF for each class is estimated using the Parzen's estimator [24] as follows

$$\mathrm{PDF}_c = \frac{1}{(2\pi\sigma_c^2)^{L/2} N_c} \sum_{i=1}^{N_c} \exp\left(\frac{-1}{2\sigma_c^2}||\lambda - \lambda_{tr_i}^c||^2\right) \quad (3)$$

Where, L is the λ_{tr} vector length, and σ_c is a smoothing factor.

The PCA weight matrix, W_{tr}, is divided into N_{cl} matrices, where each matrix contains the weights of one class. Define those matrices as $W_{tr}^{\ c}$, where c is the class index. Each matrix has a size of N_{Etr}.N_c. Define each column vector of $W_{tr}^{\ c}$ as $W_{tri}^{\ c}$, where i = 1, …, N_c. Another PDF, for each class, can be estimated as

$$\mathrm{PDF}_c = \frac{1}{(2\pi\sigma_c^2)^{L_W/2} N_c} \sum_{i=1}^{N_c} \exp\left(\frac{-1}{2\sigma_c^2}||W - W_{tr_i}^c||^2\right) \quad (4)$$

Similarly, the PCA weights matrix relating the unclassified pattern to each class, W_p, is divided into N_{cl} matrices, defined as $W_p^{\ c}$, where c is the class index, and p is the new pattern to be classified. Each column vector is defined as $W_{pi}^{\ c}$, where, i = 1, …, N_c. The summation layer of the PNN calculates the probability that the unclassified pattern belongs to each class. This is done as follows. The distance between each training pattern in each class and the unclassified pattern is found as

$$D_{p_i}^c = ||W_{p_i}^c - W_{tr_i}^c||^2 \quad (5)$$

The probability that the unclassified pattern belongs to a class c is calculated as

$$g_p^c = \frac{1}{(2\pi\sigma_c^2)^{L/2} N_c} \sum_{i=1}^{N_c} \exp\left(\frac{-1}{2\sigma_c^2} D_{p_i}^c\right) \quad (6)$$

Where, c = 1, … , N_{cl}. The output layer of the PNN uses a decision function based on $g_p^{\ c}$ to classify the unclassified pattern. Fig. 5 outlines the steps of pattern recognition.

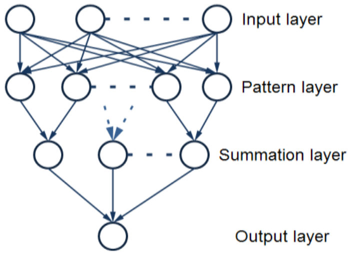


Figure 4. Illustration of the PNN layers and their connections.

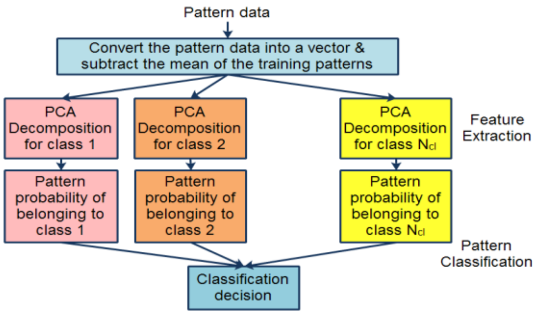

Pattern data
Convert the pattern data into a vector & subtract the mean of the training patterns
PCA Decomposition for class 1
PCA Decomposition for class 2
PCA Decomposition for class Ncl
Feature Extraction
Pattern probability of belonging to class 1
Pattern probability of belonging to class 2
Pattern probability of belonging to class Ncl
Pattern Classification
Classification decision

Figure 5. Illustration of the steps of pattern recognition.

More than one pattern maybe needed to identify the surrounding environment. This is because similar patterns can be generated from different environments, and also some environments can generate several distinctive patterns at different times. The pattern recognition algorithm is modified to avoid misclassification of the environment. Each environment has its own class, and each class consists of N_{sc} subclasses, where each subclass defines patterns with distinctive features. Training patterns are collected for each subclass. In the classification step, instead of collecting one pattern to identify the environment, N_{pe} patterns are collected over separate times. Each pattern is classified separately, and then the class that most of the patterns are classified to it is chosen, and its corresponding environment is identified as the current surrounding environment. Fig. 6 outlines this approach of environment classification. Another approach to identify the environment is to collect statistics from the unclassified patterns and include them in the classification process. On average, an urban environment has higher number of reflected signals than a suburban environment, and indoor signals are weaker and can incur longer blockage times.

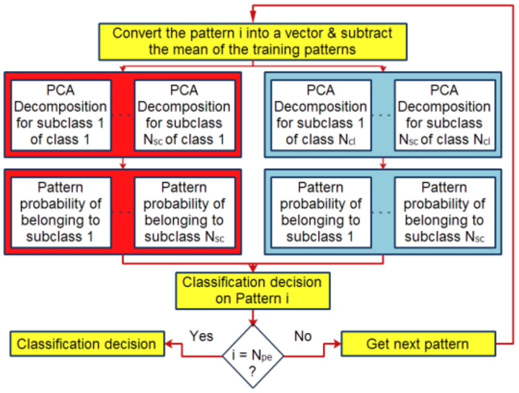

Convert the pattern i into a vector & subtract the mean of the training patterns
PCA Decomposition for subclass 1 of class 1
PCA Decomposition for subclass Nsc of class 1
PCA Decomposition for subclass 1 of class Ncl
PCA Decomposition for subclass Nsc of class Ncl
Pattern probability of belonging to subclass 1
Pattern probability of belonging to subclass Nsc
Pattern probability of belonging to subclass 1
Pattern probability of belonging to subclass Nsc
Classification decision on Pattern i
Classification decision
Yes
i = Npe ?
No
Get next pattern

Figure 6. Illustration of the steps of environment recognition.

IV. Adaptive Tracking Strategy Selection

Signal acquisition can be done under strong and weak signal conditions, and in low and high dynamics environment [3]. Following a signal acquisition, the receiver obtains rough estimates of the code delay, the Doppler shift, and the Doppler rate. A closed-loop tracking is initialized using the acquisition output, and then it works to refine the parameters estimates and continuously track changes in the code and carrier parameters. Closed-loop tracking techniques are based on generating error signals proportional to the current errors in the estimated parameters. Those error signals are fed back to the tracking algorithm to readjust the estimated parameters and maintain locks on the code and carrier parameters. A sudden large error in the estimated parameters due to sudden changes in the signal dynamics or the signal condition can cause loss of lock on the signal, and the tracking can no longer continue its operation. In this case, an acquisition process has to be re-initiated to reacquire the signal. If the signal is in an unstable environment, like continuous changes in dynamics, frequent signal blocking, or dense multipath environment, then closed-loop tracking will not be able to achieve lock on the signal. An open-loop tracking can provide the solution to tracking in such cases.

Open-loop tracking performs an acquisition-like process on the received signal, but with a smaller search range in the code delay and Doppler shift. The search range is adjusted based on the signal dynamics and condition. For example, in dense multipath reflected signals, high dynamics will require larger search range than low dynamics.

The output of the introduced MPR algorithm is fed to a tracking strategy selector, which uses the output to decide on a tracking strategy and tune the filters parameters to best suit the signal condition. The tracking algorithms previously introduced for weak signals [3], weak interrupted signals [10], and multipath tracking [14] are used as basis for closed loop tracking. The acquisition algorithms previously introduced for weak signals and high dynamics in [3] are used as basis for open loop tracking. An extra module is added to each algorithm to detect changes in the environment by detecting changes in the carrier to noise ratio C/N_0, the signal dynamics, the number of detected multipath signals, or loss of lock. Detecting changes in the signal conditions will invoke a rerun of the MPR algorithm to identify the new signal pattern and/or the environment.

V. Testing And Results

Real and simulated GPS C/A signals are used to assess the performance of the algorithm. Urban and suburban multipath patterns are generated using the models in [4, 5, 6, 7, 8] and the software in [33], while indoor and outdoor patterns are simulated. The real GPS data were provided by the University of New South Wales, Australia. The real GPS data had strong LOS signal and no multipath signals. It is processed to add few multipath non-varying signals, and then it is used to generate some of the training patterns for strong outdoor signals. Hardware and software simulators are used as sources for the simulated GPS signals. The hardware simulator's GPS data were provided by the PLAN group at the University of Calgary, Canada. The software simulator's GPS data are generated as in [3, 14]. The simulated GPS signals are used with the multipath patterns to get a variety of signals conditions to be used in the verification process.

Pattern recognition depends on the set of patterns used to train the algorithm. The classes are chosen to serve the underlying target application. The parameters chosen to identify the patterns should be enough to describe each class, and should express the differences between the classes. The target application in this paper is adaptive tracking, and so the first test is setup to serve that application. The classes are chosen to accommodate various tracking strategies. Six classes are selected for the test. The classes are as follows:

1) SL-CM: Slow varying LOS signal power with very few non-varying multipath reflected signals.

2) SL-DM: Slow varying LOS signal power with very dense multipath reflected signals.

3) FL-VM: Fast changing LOS signal power with various scenarios for multipath reflected signals.

4) BL-VM: LOS signal with occasional blockage, and with various scenarios for multipath reflected signals.

5) BL-DM: Blocked or very weak LOS signal with dense multipath reflected signals.

6) BL-LM: Blocked or very weak LOS signal with few multipath reflected signals.

A total of 150 training patterns are generated. The integration time, T_i, used here is 5 milliseconds, and each pattern spans 1 second. The patterns are divided into the six aforementioned classes, where each pattern is allocated to the most appropriate class. To test the classification performance, 25 new patterns are generated, where none of them were used in the training process. The probability that each unclassified pattern, p, belongs to one of the classes is calculated as in (6). Those probabilities are normalized as follows:

$$P_p^c = \frac{g_p^c}{\sum_{i=1}^{N_{cl}} g_p^c} \tag{7}$$

Tables I and II show 11 of the classifications results. Each column contains the result for one unclassified pattern, and each row contains the results for each class type. The pattern with the slow varying LOS signal with non-varying few multipath signals was classified with probability 1 to its class, SL-CM, which means it did not show any similarities to any of the other classes. The number and density of multipath reflected signals can vary from very light to very dense, and there is no actual boundary to classify any possible multipath density pattern into only two classes, However, the classification results indicate how light or how dense a pattern is. This is clear in the results in the last four columns of table II. This indicates that the probability of each class can be used to draw further conclusions about a pattern.

Environment identification test is also conducted. Four environments are defined: Outdoor, indoor, urban, and suburban. The outdoor is defined here as an environment that has very few surrounding objects. Urban and suburban environments structures differ from one place to another. In this test, a suburban environment is a one with wide streets, trees, and low-rise buildings with large separation between them. Urban environment is a one with less wide streets, dense and high-rise buildings. Training patterns are recorded for each environment and divided into subclasses, where the patterns for each subclass are chosen to have distinctive features relative to each other.

The outdoor environment is assigned two subclasses, where one subclass has strong LOS signal and no multipath signals, and the second subclass has strong LOS signal and few non-varying multipath signals.

The indoor environment is assigned two subclasses, where one subclass has light multipath signals and the other has dense multipath signals. Those two subclasses are characterized by weak power for all the signals.

TABLE I. PATTERN RECOGNITION CLASSIFICATION RESULTS

	Unclassified Pattern Type				
Class	**SL-CM**	**SL-DM**	**SL-DM**	**FL-VM**	**FL-VM**
SL-CM	1.0	0.017	0.015	0.009	0.012
SL-DM	0.0	0.755	0.762	0.146	0.146
FL-VM	0.0	0.193	0.189	0.648	0.687
BL-VM	0.0	0.035	0.034	0.142	0.11
BL-DM	0.0	0.0	0.0	0.03	0.008
BL-LM	0.0	0.0	0.0	0.025	0.006

TABLE II. PATTERN RECOGNITION CLASSIFICATION RESULTS

	Unclassified Pattern Type					
Class	**BL-VM**	**BL-VM**	**BL-DM**	**BL-DM**	**BL-LM**	**BL-LM**
SL-CM	0.012	0.012	0.0	0.0	0.0	0.0
SL-DM	0.254	0.269	0.0	0.0	0.0	0.0
FL-VM	0.15	0.14	0.082	0.093	0.074	0.069
BL-VM	0.564	0.562	0.033	0.039	0.029	0.028
BL-DM	0.011	0.009	0.542	0.524	0.38	0.322
BL-LM	0.009	0.007	0.343	0.344	0.517	0.437

The urban environment is assigned four subclasses as follows: (1) LOS signal with dense multipath signals; (2) LOS signal with light multipath signals; (3) blocked LOS signal with dense multipath signals, and (4) occasionally blocked LOS signal with occasionally disappearing multipath signals.

The suburban environment is assigned six subclasses as follows: (1) a subclass that characterizes the effect of trees; (2) LOS signal with light multipath signals; (3) LOS signal with dense multipath signals; (4) blocked LOS signal with light multipath signals; (5) blocked LOS signal with dense multipath signals; and (6) LOS signal with no multipath signals.

A total of 500 patterns are used for training. Twelve sets of unclassified patterns are tested, where each set had N_{pe} = 5 patterns. Each set represents patterns of one environment, and each environment type had 3 sets of patterns; define this number of sets as N_{env} (N_{env}=3).

Table III shows a summary of the results. This summary is calculated by averaging the results obtained from each environment's three sets of patterns.

TABLE III. SUMMARY OF ENVIRONMENT CLASSIFICATION RESULTS

Environment	**Outdoor**	**Indoor**	**Urban**	**Suburban**
P_{sc}	1	0.95	0.8	0.75
P_{cl}	1	1	0.95	0.9
Classification percent	100	100	100	100

The first row of the results in table III is the average of the maximum probability that a subclass has generated at each of the N_{pe} patterns, and each of the N_{env} sets, i.e.,

$$P_{sc} = \frac{1}{N_{env}} \sum_{j=1}^{N_{env}} \left\{ \frac{1}{N_{pe}} \sum_{i=1}^{N_{pe}} \max \left\{ P_p^c \right\}_i \right\}_j \tag{8}$$

The second row of the results is the average of the summation of P_p^c in (7) of the subclasses that belong to the correct environment, i.e.,

$$P_{cl} = \frac{1}{N_{env}} \sum_{j=1}^{N_{env}} \left\{ \frac{1}{N_{pe}} \sum_{i=1}^{N_{pe}} \left\{ \sum_{c=1}^{N_{sc}} P_p^c \right\}_i \right\}_j \qquad (9)$$

The third row is the percent of correct classification taken over the N_{pe} patterns of each of the N_{env} sets. As shown, the results of this test generated 100 percent correct environment classification.

The definition of environments is application dependent. The selection of subclasses depends on the structure of the selected environments. The MPR algorithm is very flexible in that the training patterns can be changed to suit the desired classification.

VI. Conclusions

This paper introduced a novel machine-learning pattern-recognition algorithm to identify the surrounding environment from the characteristics of the multipath reflected signals. The algorithm employed feature extraction and pattern classification functionalities to identify the distinctive features of the training patterns and classify new unclassified patterns into predefined classes. The predefined classes can be chosen based on the desired classification criteria. The algorithm has the flexibility to work with new classes. New environments or signal conditions can be added by simply adding new training patterns from those environments. Testing results indicated the ability of the algorithm to correctly classify multipath patterns and reliably identify the surrounding environment. This algorithm opens the door for the implementation of adaptive tracking techniques that adjust their tracking strategy based on the surrounding environment or the signal condition. Adaptive tracking techniques can play a major role in providing highly accurate and reliable positioning in wireless and mobile applications.

Acknowledgment

The author thanks Professor Gerard Lachapelle, Aiden Morrison, and the PLAN group of the University of Calgary, Canada, for providing some of the GPS simulated data used in this paper.

The author thanks Professor Andrew Dempster of the University of New South Wales, Australia, for providing the real GPS data used in this paper.

The author thanks Dr. Andreas Lehner and Dr. Alexander Steingass for their multipath channel models generator, which was used in this paper.

References

[1] B. Parkinson and J. Spilker, Global Positioning System:Theory and Applications. AIAA, 1996.

[2] P. Misra and P Enge, Global Positioning System: Signals, Measurements and Performance. Ganga-Jamuna Press, 2011.

[3] N. I. Ziedan, GNSS Receivers for Weak Signals. Artech House, Norwood,MA, July, 2006.

[4] A. Lehner and A. Steingass, "Measuring the Navigation Multipath Channel: A Statistical Analysis," In ION GNSS 2004, Long Beach, California, USA, September 2004.

[5] A. Lehner and A. Steingass, "A Novel Channel Model for Land Mobile Satellite Navigation," In ION GNSS 2005, Long Beach, CA, USA, September 13–16 2005.

[6] A. Lehner and A. Steingass, "Differences in Multipath Propagation between Urban and Suburban Environments," In ION GNSS 2008, September 2008.

[7] F. M. Schubert, A. Lehner, A. Steingass, P. Robertson, B. H. Fleury, and R. Prieto-Cerdeira, "Modeling the GNSS Rural Radio Channel: Wave Propagation Effects caused by Trees and Alleys," In ION GNSS 2009, September 2009.

[8] A. Lehner, A. Steingass, and F. Schubert, "A Location and Movement Dependent GNSS Multipath Error Model for Pedestrian Applications," In ION GNSS 2009, September 2009.

[9] M. L. Psiaki and H. Jung, "Extended Kalman Filter Methods for Tracking Weak GPS Signals," In Proc.ION GPS, Portland, OR, September 24–27, 2002.

[10] N. I Ziedan, "Extended Kalman Filter-Based Deeply Integrated GNSS/Low-Cost INS for Reliable Navigation under GNSS Weak Interrupted Signal Conditions," In ION GNSS 2006, pp. 2889–2990, Fort Worth, TX, US, September 26–29 2006.

[11] F. Van Graas, A. Soloviev, M. Uijt de Haag, and S. Gunawardena, "Closed-Loop Sequential Signal Processing and Open-Loop Batch Processing Approaches for GNSS Receiver Design," IEEE Journal of Selected Topics in Signal Processing, Vol. 3, No. 4, pp.571–586, August 2009.

[12] A. Giremus, J. Tourneret, and V. Calmettes, "A Particle Filtering Approach for Joint Detection/Estimation of Multipath Effects on GPS Measurements," IEEE Trans. on Signal Processing, Vol/ 55, No. 4, pp. 1275–1285, April 2007.

[13] P. Closas, C. Fernandez-Prades, and J. A. Fernandez-Rubio, "A Bayesian Approach to Multipath Mitigation in GNSS Receivers," IEEE Journal of Selected Topics in Signal Processing, Vol 3, No. 4, pp.695–706, 2009.

[14] N. I. Ziedan, "Multi-Frequency Combined Processing for Direct and Multipath Signals Tracking Based on Particle Filtering," In ION GNSS 2011, pp. 1090–1101, Portland, OR, USA, September 19–23. 2011.

[15] M. A. Turk and Alex P. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, Vol. 3, No. 1, pp.71–86, 1991.

[16] M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces," In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 586–591, June 3–6, 1991.

[17] J. Zhang, Y. Yan, and M. Lades, "Face Recognition: Eigenface, Elastic Matching, and Neural Nets," Processdings of the IEEE, Vol 85, No. 9, pp. 1423–1435, September 1997.

[18] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang, "Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 26, No. 1, pp.131–137, January 2004.

[19] P. S. Sandhu, I. Kaur, A. Verma, S. Jindal, I. Kaur, and S. Kumari, "Face Recognition Using Eigen face Coefficients and Principal Component Analysis," International Journalof Electrical and Electronics Engineering, Vol. 3, No. 8, pp.498–502, 2009.

[20] J. .Ashok and D.E.G. Rajan, "Principal Component Analysis Based Image Recognition," InternationalJournal of Computer Science and Information Technologies, Vol. 1, No. 2, pp. 44–50, 2010.

[21] W. Zuo, "Bidirectional PCA with Assembled Matrix Distance mMetric for Image Recognition," IEEE Trans on Systems, Man, and Cybernetics, Vol. 36, No. 4, pp. 863-872, August 2006.

[22] H. Lu, K. N. Plataniotics, and A. N. Venetsanopoulos, "MPCA: Multilinear Principal Component Analysis of Tensor Objects," IEEE Trans. on Neural Networks, Vol. 19 No. 1, pp. 18-39, 2008.

[23] V. Christlein,"An Evaluation of Popular Copy-Move Forgery Detection Approaches," IEEE Trans. on Information Forensics and Security, Vol. 7, No. 6, pp. 1841-1854, December 2012.

[24] S. Haykin and T. K. Bhattacharya, “Modular Learning Strategy for Signal Detection in a Nonstationary Environment,” IEEE Trans. on Signal Porcessing, Vol. 45, No. 6, pp. 1619–1637, June 1997.

[25] T. K. Bhattacharya and S. Haykin, “Neural Network-Based Radar Detection for an Ocean Environment”, IEEE Trans. on Aerospace and ElectronicSystems, Vol. 33, No. 2, pp. 408–420, April 1997.

[26] Y. Zhang, L. Wu, N. Neggaz, S. Wang, and G. Wei, “Remote-Sensing Image Classification Based on an Improved Probabilistic Neural Network,” Sensors, Vol. 9, pp. 7516–7539, 2009.

[27] E. Parzen, “On Estimation of a Probability Density Function and Mode,” Annals of Mathematical Statistics, Vol. 33, No. 3: pp. 1065–1076, September 1963.

[28] D. F. Specht, “Probabilistic Neural Networks for Classification, Mapping, or Associative Memory,” In IEEE International Conference on Neural Networks, Vol. 1, pp. 525–532, July 24–27 1988.

[29] D. F. Specht, “Probabilistic Neural Networks,”Neural Networks, Vol 3, pp.109–118, 1990.

[30] D. F. Specht, “Probabilistic Neural Networks and the Polynomial Adaline as Complementary Techniques for Classification,” IEEE Trans. on Neural Networks, Vol. 2, No. 1, pp.111–121,March 1990.

[31] K. Z. Mao, K.-C. Tan, and W. Ser, “Probabilistic Neural-Network Structure Determination for Pattern Classification,” IEEE Trans. on Neural Networks, Vol. 11, No. 4, pp. 1009–1016, July 2000.

[32] L. Rutkowski, "Adaptive probabilistic neural networks for pattern classification in time-varying environment," IEEE Trans. on Neural Networks, Vol. 15, No. 4, pp. 811-827, July 2004.

[33] A. Lehner and A. Steingass. SatelliteNavigation Multipath Channel Models. http://www.kn-s.dlr.de/satnav.

AUTHOR PROFILE

Nesreen I Ziedan is an Assistant Professor at the Computer and Systems Engineering Department, Faculty of Engineering, Zagazig University, Egypt. She received a Ph.D. degree in Electrical and Computer Engineering from Purdue University, West Lafayette, IN, USA, in 2004. She also holds an M.S. degree in Control and Computer Engineering, a Diploma in Computer Networks, and a B.S. degree in Electronics and Communications Engineering. Dr. Ziedan has several U.S. Patents in GPS receivers design and processing, she is the author of a book entitled “GNSS Receivers for Weak Signals”, and she is an Associate Fellow of the Royal Institute of Navigation, UK.

Low cost approach to real-time vehicle to vehicle communication using parallel CPU and GPU processing

GOH CHIA CHIEH
Faculty of Electronic and Electrical Engineering
University of Nottingham
Selangor, Malaysia

DINO ISA
Faculty of Electronic and Electrical Engineering
University of Nottingham
Selangor, Malaysia

***Abstract*— This paper proposes a novel Vehicle to Vehicle (V2V) communication system for collision avoidance which merges four different wireless devices (GPS, Wi-Fi, ZigBee® and 3G) with a low power embedded Single Board Computer (SBC) in order to increase processing speed while maintaining a low cost. The three major technical challenges with such combinations are the limited system bandwidth, high memory requirement and slow response time during data processing when accessing various collision avoidance situations. Collision avoidance data processing includes processing data for vehicles on express ways, roads, tunnels, traffic jams and indoor V2V communication such as required in car parks. Effective methods are proposed to address these technical challenges through parallel Central Processing Unit (CPU) and Graphic Processing Unit (GPU) processing. With this, parallel V2V trilateration and parallel bandwidth optimization, multi-dimensional real time complex V2V data streaming can be attained in less than a second. The test results have shown that there is at least a 4 to 10 times improvement on processing speed with parallel CPU and GPU processing used in V2V communication depending on different road safety conditions.**

Keywords-component; CUDA; Parallel processing; Vehicle to Vehicle Communication; WLAN; ZigBee.

I. INTRODUCTION

Road safety has drawn worldwide attention due to an increasing number of accidents every year. According to the United Nations (UN) road safety report approximately 90% of road fatalities occur in low and middle income countries [1] and high portion of these accidents are due to issues directly related to the driver.

Due to its high impact, much research on Vehicular Ad-Hoc Networks (VANET) and the safety applications that go along with it have been carried out to prevent or lower the rate of accidents [2] [3] [4] [5]. VANETs are the customary implementation of network communications for Intelligent Transport Systems (ITS). Dedicated Short-Range Communication (DSRC) is the standard used for V2V communication operating in 5.9 GHz. The most basic V2V communication system consists of GPS and General Packet Radio Service (GPRS) which transmit the current vehicle location to the nearby vehicle using mobile the Point to Point (P2P) network protocol.

VANET is designed to target for quicker response of the low latency network V2V communication connectivity. However as the infrastructure of the vehicles are getting more complex and the flow of traffic is increased, more sensors and electronic devices are built on-board the vehicle to enhance road safety conditions. Much research and the accompanying solutions have been implemented to lower down the cost of replacing the wireless V2V communication system which uses expensive transceivers with commonly used wireless network systems such as Wi-Fi (802.11), Wireless Broadband (802.16) ZigBee® (802.15.4) or other emerging network device [6] [7] [8] [9] [10]

The objectives of this research are as follows:

- LOW COST -To design a low cost V2V collision avoidance system through parallel CPU and GPU processing based on widely available GPS, Wi Fi, ZigBee® and 3G broadband units.
- LOW POWER, HIGH SPEED - Development of parallel bandwidth, data optimization and V2V trilateration algorithm, able to process the data in real time on a low power SBC.

II. RELATED WORK

Vehicular ad-hoc networks (VANET) have gain wide popularity since the mid-90s [11]. VANET is an important element for accident prevention for on road vehicle. Data retrieval and clear presentation to the driver from the surrounding environment has been proven to reduce human errors while driving [12] [13]. However, the wireless bandwidth for vehicle communication system is very limited [14] [15]. An efficient communication protocol is needed to avoid overloading the system.

In view of this, [16] [17] [18] [19] et. al proposed the GPS and GPRS/GSM communication which tracks the location of the vehicle. The data is uploaded to the server and is monitored in real time. The main weakness for the system is that, the data is transmitted or received through GPRS/GSM which only allows a small amount of data to be transferred in and out the system.

[20] [21] [22] [23] et. al proposed the V2V communication by implementing the Collision Warning Systems (CWS) which

are radar, camera and radio based. Radar/Ladar and camera-based collision detection is meant for detecting and viewing what is in front and at the rear of the vehicle. Their proposed method uses a radio based system which covers the side angle where a potential collision might occur. Radio based methods adopt a similar concept such as Wi-Fi stack - Time Division Multiple Access (TDMA). The main drawback for such setup is the system cost increases due to the implementation of Radar.

[24] [25] [26] [27] et. al proposed the Wireless Access for Vehicular Environment (WAVE) for V2V and Vehicle to Infrastructure (V2I) system. The proposed system uses the Road Side Unit (RSU) coordination scheme for WAVE safety services support. This means that, the Access Point (AP) or known as WAVE provider is set up along the road. The economic drawback of this method is the need to set up the high cost RSU's and this limits the number of locations that is covered.

Our work is focused on creating a cost effective real time V2V network system. This system in each vehicle consists of a Global Positioning System (GPS) satellite based positioning system and a wireless location positioning system (LPS) which was developed using low cost short range wireless ZigBee® device (IEEE 802.15.4). 3G broadband is used to transmit or receive the information from the server in case of emergencies such as vehicle breakdown, traffic jam or road accidents. The methodologies for such configurations are implemented through parallel CPU and GPU processing. This method is capable of not only boosting the performance occupancy but also reduces the delay latency on data transfer between hardware and software in an efficient order.

III. PROPOSED WORK

As shown in Fig. 1, is a top down embedded system architecture view. It consists of 5 layers, Red, Green, Blue, Purple and Orange. The system can be categorized into two parts: real time data processing and non-real time data processing. The real time requirement for data processing such as with the Wi Fi, ZigBee® and GPS will be given a higher priority as compare to Broadband transmission and speech recognition. Non real time data processing

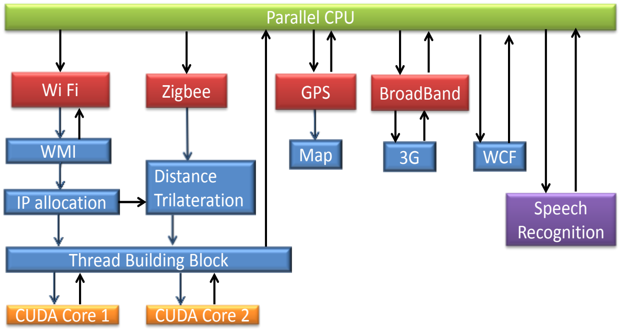


Figure 1. System Architecture

A. *The Red and Green Layer*

This is the hardware layer where Wi Fi, ZigBee®, GPS and Broadband Modem (represented in the block diagram) are connected and data is obtained from the real world. Data synchronization between devices was previously performed at the Green layer - Task parallelism library (TPL). TPL is one of the features in Microsoft® .net framework. The objective of using TPL is to divide the task into sections and parallel processed through the CPU. The incoming data from different devices were split into four different tasks. Each task was split into multiple threads. Threads can be parallel processed through multiple Central Processing Unit (CPU) and Graphic Processor Unit (GPU). If different data streams from the hardware devices simultaneously, data processing priority was given from the left to the right, which means the Wi Fi device has the highest priority and the Broadband device has the lowest.

Both Wi Fi device and ZigBee® consist of two parts, with a total of four devices. There are two devices in Wi Fi block diagram: Wi Fi router and Wi Fi client adapter. The Wi Fi client was installed on individual vehicle. Data transfer between Wi Fi client adapters of different vehicle is facilitated through the Wi Fi router. The maximum transmission range for the Wi Fi device is 100 meter radius wide, supporting up to 54Mbps of data transfer rate.

ZigBee® system consists of two items: the ZigBee® Coordinator and the ZigBee® router. The ZigBee® router is used to establish connection between vehicles in a mesh topology the total number depending on the ZigBee® system standard for number of devices within a mesh. Although the maximum data transfer rate for ZigBee® device is 250Kbps, it has a very low packet overhead during transfer. The data transfer rate will maintain its consistency within 100 meter radius. ZigBee® devices were used in V2V communication mainly for range estimation. ZigBee® devices are very low in power consumption. Hence, the device will therefore continue to operate even when the vehicle engine is turned off.

For outdoor location tracking, the Global Positioning System (GPS) is used. The GPS device follows the NMEA 0183 format for location and time zone retrieval. The fundamental problems for GPS are: the distance estimation tolerance is up to ±15 meters [37] and GPS can only work in outdoor environment. Therefore, the combination of ZigBee® and GPS is a necessity to solve the problem.

The 3G broadband device is used for internet connection to send or download the latest update from the centralized management server. This information includes traffic info and accident location from the vehicle itself (if any).

B. *The Blue Layer*

The software layer, represented in Blue Layer, is the interface to the Hardware layer (Red Layer). Vehicle to vehicle communication was facilitated using Windows Communication Foundation (WCF) framework, which is a replacement for Winsock API.

Windows Communication Foundation served as the Transmission Control Protocol/Internet Protocol (TCP/IP) stack to support communication with the router. Each vehicle has its designated Internet Protocol (IP) address and a unique ID tag. The ID tag can be transmitted through ZigBee® device from vehicle to vehicle in a star topology that used multicast protocol. The IP address is generated in random order; it follows the IPV4 format, Class C network ranging from IP address 192.168.9.4 to 192.168.9.254. IP address which range from 192.168.9.1 to 192.168.9.3 is reserved for debugging and testing purposes. If IP conflict occurs, Windows Management Instrument (WMI) will assign a new IP to the current vehicle. The vehicle to vehicle communication is within 100 meters radius and has maximum of 30 vehicles linked within the range. Each time a communication is established, the router search through the network IP polls (192.168.9.4 to192.168.9.254) for unique ID. Once the ID was identified, the properties of the vehicles, including the speed, location and the ID, are sent through the wireless communication. Once the vehicle is out of the 100 meter radius range, the IP will be released automatically so that the IP address will be available for another incoming nearby vehicle.

Without GPS, the location of the vehicle is calculated by using the distance trilateration method. This method pinpoints the vehicle location by comparing the distance of the 3 nearby vehicle. Once the location of the nearby vehicle location is identified, the vehicle location is computed through the Thread Building Block (TBB) [39]. TBB allows the thread to be parallel processed on two different GPU cores. Whenever the GPS signal is available, the map will be updated accordingly to reflect the latest location of any vehicle within the 100 meters vicinity radius range.

The vehicle ID, location, speed, time and date will be sent to the server through 3G broad. The server will help to monitor the traffic flow. If any accident happened in the particular location, the server will transmit the data to alert the driver.

C. The Purple Layer

Speech recognition was located at the purple layer. It was integrated into the system based on the built-in function from Windows OS. The system is programmed to detect extreme changes in voice pitch, e.g. shouting. The system has some pre-programmed command such as “send it”, “retrieve it” that allows user to interact with the system while driving. Once the system detects an abnormal tone, the system will switch to panic mode and take control of the dangerous situation automatically. The system will give audible warning such as “warning, nearby vehicle is within the range of xx meters”. Information regarding the vehicle, for example the location and the vehicle ID, will be transferred to the server as needed.

D. The Orange Layer

The orange layer represents the Graphic Processing Unit (GPU) parallelism layer. Compute Unified Device Architecture (CUDA) programming model has two parts: Host and Device. The Host is the CPU and the Device is the GPU. Data were transferred from the Host (CPU) to the Device (GPU) in a single-program, multiple data (SPMD) for preprocessing. Thread building block [39] transfers the array of data from the parallel CPU to the CUDA core 1 and CUDA core 2 for parallel processing. The device will send the final results back to the Host. Standard C++ compiler is used for compiling data in the host, while CUDA’s own ANSI C based compiler is used for compiling in the device (GPU) coding.

IV. Methodology

A. Received Signal Strength Indicator (RSSI) Trilateration to find vehicle location.

Trilateration was used to locate a vehicle in the vicinity of three other vehicles. This was done in order to monitor the relative movement of vehicles surrounding each other to predict occurrences of abrupt evasive action such as when blind spots are encountered which leads to dangerous situations. This was accomplished by monitoring wireless signal strength. GPS was not used in this case because it cannot be used in indoor (car park) or in a sheltered operation.

The basic concept of wireless signal strength is shown Fig. 2. There are two wireless devices, as the two devices getting further apart, signal strength within the radius will be getting weaker. The measurement of the signal strength which received from the device is known as Received Signal Strength Indicator (RSSI). RSSI measures the power which received from the antenna in decibel meter (dBm). Thus, higher positive RSSI value indicators stronger signal strength.

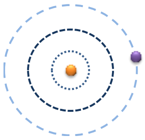


Figure 2. Received Signal Strength

Based on the signal strength, there are 3 common methods to measure the distance between two points: Received Signal Strength Indicator (RSSI), Time of Arrival (TOA) and Angle Of Arrival (AOA). The simplest and cheapest method is the Received Signal Strength (RSS) method as it doesn’t required additional hardware.

As shown in Fig. 3, in order to estimate a location, at least three wireless devices were needed to triangulate the position of the 4th wireless device. There are three wireless devices, a, b and c. Wireless device d is within the intersection of a, b and c. The coordination of a, b and c can be obtained from either the Wi Fi or a fix point ZigBee® router from the vehicle.

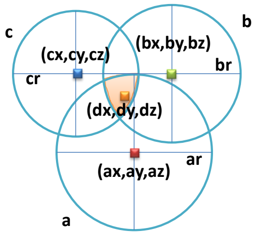


Figure 3. RSSI trilateration

As shown in Fig. 3 above, the equation of the trilateration can be written as:

$$\begin{bmatrix}(x_a-x_d)^2+(y_a-y_d)^2+(z_a-z_d)^2\\(x_b-x_d)^2+(y_b-y_d)^2+(z_b-z_d)^2\\(x_c-x_d)^2+(y_c-y_d)^2+(z_c-z_d)^2\end{bmatrix}=\begin{bmatrix}r_a{}^2\\r_b{}^2\\r_c{}^2\end{bmatrix} \qquad (1)$$

Where

r_a, r_b, r_c = radius of circle a, b and c

x_a, x_b, x_c, y_a, y_b, y_c, z_a, z_b, z_c = x, y and z of GPS coordination of circle a, b, c and d

The intersection point of these three circles can be anywhere in the shaded orange area. In order to minimize the estimated error, the least square error method can be used:

$$\begin{bmatrix}[x_a-(x_d+Wd_x)]^2+[(y_a-(y_d+Wd_y)]^2+[z_a-(z_d+Wd_z)]^2\\ [x_b-(x_d+Wd_x)]^2+[(y_b-(y_d+Wd_y)]^2+[z_b-(z_d+Wd_z)]^2\\ [x_c-(x_d+Wd_x)]^2+[(y_c-(y_d+Wd_y)]^2+[z_c-(z_d+Wd_z)]^2\end{bmatrix} \qquad (2)$$

$$-\begin{bmatrix}r_a{}^2\\r_b{}^2\\r_c{}^2\end{bmatrix}=\begin{bmatrix}E1^2\\E2^2\\E3^2\end{bmatrix}$$

Square Error = $E1^2 + E2^2 + E3^2$ (3)

Where

Wd_x,Wd_y, Wd_z = weight for dx, dy, dz coordination

From equation (2), the equation can be expanded as:

$$\begin{aligned}A=&x_a{}^2-2\left[x_a\left(x_d+Wd_x\right)\right]+\left(x_d+Wd_x\right)^2+\\&y_a{}^2-2\left[y_a\left(y_d+Wd_y\right)\right]+\left(y_d+Wd_y\right)^2+\\&z_a{}^2-2\left[z_a\left(z_d+Wd_z\right)\right]+\left(z_d+Wd_z\right)^2-r_a{}^2\end{aligned} \qquad (4)$$

$$\begin{aligned}B=&x_b{}^2-2\left[x_b\left(x_d+Wd_x\right)\right]+\left(x_d+Wd_x\right)^2+\\&y_b{}^2-2\left[y_b\left(y_d+Wd_y\right)\right]+\left(y_d+Wd_y\right)^2+\\&z_b{}^2-2\left[z_b\left(z_d+Wd_z\right)\right]+\left(z_d+Wd_z\right)^2-r_b{}^2\end{aligned} \qquad (5)$$

$$\begin{aligned}C=&x_c{}^2-2\left[x_c\left(x_d+Wd_x\right)\right]+\left(x_d+Wd_x\right)^2+\\&y_c{}^2-2\left[y_c\left(y_d+Wd_y\right)\right]+\left(y_d+Wd_y\right)^2+\\&z_c{}^2-2\left[z_c\left(z_d+Wd_z\right)\right]+\left(z_d+Wd_z\right)^2-r_c{}^2\end{aligned} \qquad (6)$$

By using Newton-Raphson method to find the trilateration of circle a, b and c:

$$\text{F }=\begin{bmatrix}A(x,y,z)\\B(x,y,z)\\C(x,y,z)\end{bmatrix} \qquad \Delta=\begin{bmatrix}\Delta x\\\Delta y\\\Delta z\end{bmatrix}$$

$$F'=\begin{bmatrix}\frac{\partial A(x,y,z)}{\partial x}&\frac{\partial B(x,y,z)}{\partial y}&\frac{\partial C(x,y,z)}{\partial z}\\\frac{\partial A(x,y,z)}{\partial x}&\frac{\partial B(x,y,z)}{\partial y}&\frac{\partial C(x,y,z)}{\partial z}\\\frac{\partial A(x,y,z)}{\partial x}&\frac{\partial B(x,y,z)}{\partial y}&\frac{\partial C(x,y,z)}{\partial z}\end{bmatrix}$$

Where

$$F+F^{'}\Delta=0 \qquad \Delta=-\frac{F}{F'}$$

$$\Delta x=-\frac{F(x)}{F'(x)} \quad \Delta y=-\frac{F(y)}{F'(y)} \quad \Delta z=-\frac{F(z)}{F'(z)}$$

$$x_{n+1}=x+\Delta x \quad y_{n+1}=y+\Delta y \quad z_{n+1}=z+\Delta z$$

The initial root value x_{n+1}, y_{n+1}, z_{n+1} of function F is chosen arbitrarily at the beginning. Every *n* iteration, check the absolute value of gx, gy and gz:

$$gx=ABS\left[-\frac{F(x)}{F'(x)}\right] \quad gy=ABS\left[-\frac{F(x)}{F'(x)}\right] \quad gz=ABS\left[-\frac{F(x)}{F'(x)}\right]$$

The algorithm will stop at *n* iteration if gx, gy and gz value ≤ 0.1. As gx, gy and gz get closer to 0, the estimated vehicle location gets closer to the coordinate of dx, dy and dz.

B. Bandwidth Optimization

1) Media Access Control (MAC) Address Lookup Table (LUT)

In order to implement the technique discussed previously, the ZigBee® was used to initiative data transfer from vehicle to vehicle and the WiFi network was used to transfer higher volumes of data. The Zigbee® has a lower overhead and therefore establishes the connection faster while the WiFi has a larger overhead (slower connection) but is able to handle higher amounts of data. Hence by using both in parallel we obtain a better performance.

Bandwidth optimization for the ZigBee® and WiFi systems is important because of the amount of data transfer which occurs from the surrounding vehicles at any given time. This amount adds uncertainty to the system hence optimization is needed. As this information is also processed in real time to predict dangerous situations, bandwidth optimization is a necessary part of feasibly implementing the trilateration technique discussed previously. In order to optimize the bandwidth usage for the Wi Fi, a Lookup Table (LUT) is used to map the MAC addresses of the Wi Fi router with ZigBee® coordinator and ZigBee® router.

As previously mentioned the WiFi network is used to transfer trilateration data between vehicles. Wi Fi was designed for data streaming and the maximum data streaming bandwidth from one Wi Fi device to another is up to 54 Mega bit per second (Mbps). However, establishing a connection between a Wi Fi client and a Wi Fi router connection requires some time.

As shown in Fig. 4, OpenWRT Wi Fi router is configured to be in a mesh topology. Wi Fi routers (yellow dot) can take up to 5 second to establish a single connection from the client (purple dots). It is more common to have the Wi Fi router to be fixed permanently on a particular spot, for example inside a building whereby the clients join the same Wi Fi router network.

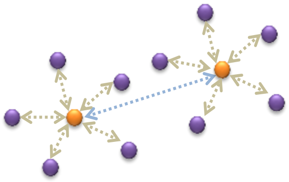


Figure 4. Wi Fi router and Wi Fi Clients

In this V2V communication system, there are 5 devices: GPS, ZigBee® coordinator, ZigBee® router, Wi Fi client and Wi Fi router. Each device of every vehicle has a unique MAC address and the same MAC address cannot be repeated.

Based on the Wi Fi 802.11 standards [40] Wi Fi MAC address can be obtained whenever the Wi Fi wireless signal is transmitted. The signal contains information on the Network Type, Authentication, Encryption and Basic Set Service Set Identification (BSSID).

BBSID has divided into 5 sections: wireless MAC address, Signal Strength, Signal Type and Data Transfer rate represented in Mega bit per second (Mbps).

The Wi Fi router wireless MAC address format is as follow:

CC: XX: XX: XX: XX: XX

CC = country code from 00- FF (Hexadecimal), representing 256 different countries based on a unique country code ID store on the server.

XX = Wi Fi MAC address (expressed in 2 bytes Hexadecimal from 00 – FF)

Since the Wi Fi router's MAC address is unique for each vehicle, the following LUT is generated:

CC: XX: XX: XX: XX: XX, YY: YY: YY: YY: YY: YY: YY: YY, ZZ: ZZ: ZZ: ZZ: ZZ: ZZ: ZZ: ZZ

Where

YY= ZigBee® coordinator MAC address

ZZ= ZigBee® router (node) MAC address

2) Data Transfer Optimization

The wireless distance from one device to another is directly proportional to the rate of change of data transfer. In other words, the further the distance between vehicle 1 and vehicle 2 is, the weaker the wireless signal will be. Weak wireless signal limit less data to be transmitted or received from vehicle 1 to vehicle 2.

In order to find the optimal data transmission distance between vehicle 1 and vehicle 2, data optimization method is implemented to the system. As shown in Table I, the data transmission occurs only when the Wi Fi signal strength is greater than 10%. If the signal is low, the data which has the highest priority will be transmitted.

TABLE I. WI FI SIGNAL STRENGTH

Wi Fi Signal Strength (%)	Types of data to be transmitted
Less than 10%	X
10% to 29%	1
20% to 39%	2
40% to 59%	3
60% to 79%	4
80% to 100%	5

The data transfers are as follows:

1. Transmit the GPS coordination for maximum of 5 vehicle within 20 meters Wi Fi range
2. Scan for MAC address of ZigBee® coordinator and router of any vehicle nearby, once it found, linked up all the nodes together so that data can be transferred from one vehicle to another. Scanning stop when it reaches 10 vehicles or GPS coordination with Wi Fi radius of 40 meters (whichever comes first).
3. Scan and established the coordination between Wi Fi and ZigBee®, for any nearby vehicle. Stop scanning after it reaches 20 vehicles or Wi Fi range of 60 meters (whichever comes first)
4. Scan and established the connection of up to 25 vehicles up to 80 meters Wi Fi radius.
   Scan and established the connection of up to 30 vehicles up to 100 meters Wi Fi radius

C. Parallel CPU and GPU Implementation

The Parallel Central Processing Unit (CPU) and Graphic Processing Unit (GPU) improve overall system performance with greater processing speed. It is crucial to implement the parallelization threads in a proper manner. In order to obtain an optimal result with parallelism, the system is relies strongly on the combination of CPU and GPU hardware performance strength, specifically favoring different type of memories for prioritize memories sequential access [41]. Currently, the limitation on CPU and GPU bandwidth become the bottleneck for hardware acceleration. There are limited bandwidth data to compute per cycle, CPU and GPU have to wait for the cycle to complete before proceeding. Compute Unified Device Architecture (CUDA), is a multithreaded GPU programming library designed by NVidia® to perform general purpose computation on a GPU. The GPU threads are executed through kernel [42]. A kernel has similar function to C programming function; the only difference is that the kernel function is requested from the CPU to the GPU device. The host (CPU) will sent the matrix array of data to the device (GPU) for arithmetic calculation. The final arithmetic value is then transferred back to the host.

1) Parallel CPU and GPU Implementation

The system is implemented in two folds. As shown in Fig. 1, there are 4 tasks handled by WLAN, ZigBee®, GPS and 3G broadband. Firstly prioritize the task. Among them, both WLAN and ZigBee® have the highest priority, followed by GPS and 3G broadband. Secondly, split the tasks into multiple threads. By using "divide and conquer" technique, each thread is partitioned into multiple parts. At the maximum performance occupancy, each part will not have more than 512 threads.

Lastly, the data matrix is sent to GPU for parallel calculation. Once the final answer has been obtained from the parallel GPU, it is transfer back to CPU for further processing.

Table II shows the 4 tasks implemented using Parallel CPU and GPU or Parallel CPU processing.

TABLE II. PARALLEL TASK DISTRIBUTION

Task	Parallel CPU + GPU	Parallel CPU
Task 1 (Bandwidth optimization)	✓	
Task 2 (GPS location transmission)		✓
Task 3 (3G broadband data transmission)		✓
Task 4 (vehicle location trilateration)	✓	

TABLE III. PSEUDO CODE (TASK 1 – BANDWIDTH OPTIMIZATION):

```
1. Create a task and name it as Task1 using Task parallel Library (one of the .net framework 4.0 features)
2. Search nearby vehicle Wi Fi router for BSSI wireless MAC address, if different BSSI wireless MAC address is found, go to step 3, else go to step 8.
3. Allocate the matrix array in GPU device
4. Store the WiFi MAC and LUT to global kernel of the GPU
5. Compare the WiFi MAC to match the LUT
6. Once a match is found, transfer the match back to CPU host. If not, continue with Step 6. If no match is found, go to Step 2.
7. Connect the wireless client based on the match, check for wireless data transfer optimization algorithm to optimize the data transfer bandwidth. If the WiFi router is further than 100 meters radius, release the IP for the next vehicle. If any IP address collision, reconfigure a new IP address.
8. Go back to step 2.
```

TABLE IV. PSEUDO CODE (TASK 2 – GPS LOCATION TRANSMISSION):

```
1. Create a task and name it as Task2 using Task parallel Library (one of the .net framework 4.0 features)
2. Check for CPU usage if it is more than 95%, if true, delay step 2 until progress it is less than 95%.
3. Send the NMEA code to the GPS device to get the coordination of x, y and z. Get the amount of satellites, maximum 12. At the same time get the date and time.
4. Broadcast the GPS coordination to the wireless network within the same transmission range.
5. Check for incoming data from other vehicle, if the distance is too close, hidden vehicle at winding road; activate speech audio to warn drivers.
```

TABLE V. PSEUDO CODE (TASK 3 – 3G BROADBAND DATA TRANSMISSION):

```
1. Create a task and name it as Task3 using Task parallel Library (one of the .net framework 4.0 features)
2. Check for CPU usage if it is more than 95%, if so, delay step 2 until progress it is less than 95%.
3. Check whether the 3G wireless broadband is more than 80% and is stable. If not don't transmit to the server. (Step 3 is to avoid data transmission error)
4. Check whether current vehicle stop at certain location for a certain time period at middle or side of the road. If so transmit the coordination to the server, the vehicle could be having an accident or break down on the road.
```

2) Two dimension Newton-Raphson using parallel CPU and GPU

Used in the trilateration calculation, the Newton-Raphson algorithm parallel CPU and GPU optimization steps are as follow:

Radius of circle r_a, r_b, r_c in meters = $10^{(RSSI\text{-}A+X\sigma)/10n}$ (7)

Where

A = RSSI value of distance in 1 meters = -40dBm

Xσ = Value of random noise = 0

n = path loss component value from 2 to 4.

From Equation (4), (5), (6):

Assume

Wd_x, Wd_y, Wd_z = 0

$$Jacobian\,Matrix = J = \begin{bmatrix} \frac{\partial A(x,y,z)}{\partial x_d} & \frac{\partial B(x,y,z)}{\partial y_d} & \frac{\partial C(x,y,z)}{\partial z_d} \\ \frac{\partial A(x,y,z)}{\partial x_d} & \frac{\partial B(x,y,z)}{\partial y_d} & \frac{\partial C(x,y,z)}{\partial z_d} \\ \frac{\partial A(x,y,z)}{\partial x_d} & \frac{\partial B(x,y,z)}{\partial y_d} & \frac{\partial C(x,y,z)}{\partial z_d} \end{bmatrix}$$

$$= \begin{bmatrix} -2x_a+2x_d & -2y_a+2y_d & -2z_a+2z_d \\ -2x_b+2x_d & -2y_b+2y_d & -2z_b+2z_d \\ -2x_c+2x_d & -2y_c+2y_d & -2z_c+2z_d \end{bmatrix} \tag{8}$$

$$Inverse\,matrix\,of\ J = J^{-1} = \frac{1}{Determinant\ J} * C^T$$

$$Adjoin\,of\ J = C^T = \begin{bmatrix} -2x_a+2x_d & -2x_b+2x_d & -2x_c+2x_d \\ -2y_a+2y_d & -2y_b+2y_d & -2y_c+2y_d \\ -2z_a+2z_d & -2z_b+2z_d & -2z_c+2z_d \end{bmatrix}$$

$$\begin{aligned} &Determinant\ J = \\ &-2x_a+2x_d\ [(-2y_b+2y_d)(-2z_c+2z_d)-[(-2z_b+2z_d)(-2y_c+2y_d)]] \\ &-[-2y_a+2y_d[(-2x_b+2x_d)(-2z_c+2z_d)-[(-2z_b+2z_d)(-2x_c+2x_d)]] \\ &+[-2z_a+2z_d[(-2x_b+2x_d)(-2y_c+2y_d)-[(-2x_c+2x_d)(-2y_b+2y_d)]] \end{aligned} \tag{9}$$

$$\begin{pmatrix} x_n \\ y_n \\ z_n \end{pmatrix} = \begin{pmatrix} x_d \\ y_d \\ z_d \end{pmatrix} - \left[\left(J^{-1}\right) \begin{pmatrix} A \\ B \\ C \end{pmatrix} \right] \tag{10}$$

*Note: The initial value for x_d, y_d, z_d were assigned randomly on the first iteration.

TABLE VI. PSEUDO CODE (TASK 4 – VEHICLE LOCATION TRILATERATION):

```
1. Create a task through Task parallel Library (one of the .net framework 4.0 features) and name it as Task4.
2. Within Task4 parallel function, check for the coordination of the GPS from nearby vehicle, if no GPS signal can be found, take the RSSI value only.
3. Split the Task into two threads using Thread Building Block so that it can be parallel process using two GPU core at the same time.
4. Allocate the array of 100 x 100 matrix array in GPU
5. Allocate the block and threads in GPU global memory
    a. Compute the RSSI to distance value using equation (7)
```

```
        b.  Substitute the coordinate value of $x_a, y_a, z_a, x_b, y_b, z_b, x_c, y_c, z_c$ into equation
            (4), (5) and (6)
        c.  Compute Inverse matrix of J from equation (9)
        d.  Allocate the share memory 10x10x10 of GPU:
            shared_memory[threadIdx.x][threadIdx.y][threadIdx.z]
        e.  Initial guess value for $\begin{pmatrix}x_d\\y_d\\z_d\end{pmatrix}=\begin{pmatrix}200\\200\\200\end{pmatrix}$ While
            $\begin{pmatrix}A\\B\\C\end{pmatrix}\geq\begin{pmatrix}0.1\\0.1\\0.1\end{pmatrix}$
                i.   Check weather is it the first time running the While loop, if so take
                     the initial value of $\begin{pmatrix}x_d\\y_d\\z_d\end{pmatrix}$ Step e) or
                     else skip step i and proceed straight to step ii.
                ii.  shared_memory[threadIdx.x][threadIdx.y][threadIdx.z] = Compute the $\begin{pmatrix}x_d\\y_d\\z_d\end{pmatrix}$
                     equation (10) using
                iii. synchronous threads
                iv.  move the value to $\begin{pmatrix}x_n\\y_n\\z_n\end{pmatrix}=\begin{pmatrix}x_d\\y_d\\z_d\end{pmatrix}$
6.  Transfer the final answer of $x_n$, $y_n$, $z_n$ from the GPU back to CPU
    Compute the final coordinate $x_n$, $y_n$, $z_n$ of current vehicle location and transmit to other vehicle using Windows Communication Foundation (WCF).
```

V. Results and discussion

A. Software Setup

The following software in Table VII is used to compile and implement the proposed parallel system:

TABLE VII. SOFTWARE SETUP

Software	Methodology/Algorithms used in this research
CUDA 4.1 [28]	Parallel GPU processing for WiFi and ZigBee® transmission protocol and bandwidth optimization.
Visual Studio 2010 (C# and C++) [29]	C# for handling non time critical task, that is graphical user interface (GUI). C++ for handling time critical real time task WiFi and ZigBee®, GPS and 3G broadband interface
Windows Communication Foundation 4.0 [30]	V2V WiFi IP address assignment
Microsoft® .Net Framework 4.0 [31]	Parallel CPU processing. The total work load is split into 4 tasks which consist of Bandwidth optimization, GPS location transmission, 3G broadband data transmission and vehicle location trilateration.
Windows Management Instrumentation [32]	V2V WiFi IP address conflict reset.
Speech Recognition SDK 11.0 [33]	Based on Microsoft® OS Speech recognition

B. Hardware Setup

The system was tested on a 2.6GHz x86 Intel Centrino dual core mobile processor, with 4 Gigabytes of RAM and two NVidia® 8800M GTX GPU graphic card with Scale Link Interface (SLI) disabled. The 8800GTX consist of 768MB RAM, 6 stream multiprocessors, with each multiprocessor consist of 16 stream processors, which comes to the total of 96 stream processors. Each multiprocessor has 8192 of 32-bit registers; each multiprocessor can have up to 768 threads. Threads are partitioned into thread blocks which can consist of up to a maximum of 512 threads, and can be furthered be divided into warps of 32 threads.

The following hardware devices in Table VIII used in this work are:

TABLE VIII. HARDWARE SETUP

Device	Function	Methodology/Algorithm
OpenWRT Wi-Fi WRT54g router (802.11g) [34]	V2V communication for real time location, speed and time tracking	Based on Open source Wi-Fi stack and customized to fit specifically in this research.
OpenWRT Wi-Fi WRT54g wireless client adapter (802.11 g) [34]	Client for V2V communication to OpenWRT router communication, up to maximum of 30 clients per router.	
ZigBee® Coordinator (802.15.4) [35]	ZigBee® client for V2V communication	Based on ember [36] ZigBee® stack. Based on RSSI for vehicle trilateration and Bandwidth optimization
ZigBee® Router (802.15.4) [35]	ZigBee® router to communicate with a ZigBee® coordinator	
GPS [37]	Vehicle location, time and date	NMEA 0183 GPS protocol to retrieve the current vehicle location.
3G broadband [38]	In an emergency situation, the device will send the vehicle location to database server.	standard 3G software protocol for transmitting and receiving from client to client , client to server and V2V communication.
LCD Touch Screen	Display vehicle location	-

C. Trial run

As a trial run, the system was tested in an outdoor environment on a local surface road. For safety reasons the vehicle maximum speed limit was set to 60kmh. For indoor testing, the GPS setting is deactivated. The vehicle to vehicle distance estimation is based on the ZigBee® trilateration and Wi Fi signal for distance judgment.

As shown in Fig. 5, the red spot shows the current vehicle location while the green spot indicate other vehicles within 100 meters radius. With this information, the system analyzes the potentially dangerous incoming vehicles and warns the driver about vehicles at blind spot, hidden corners or idle vehicle on the middle of the road.

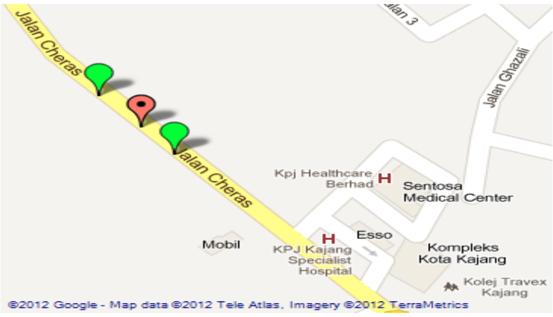


Figure 5. Map and vehicle info

Table IX shows the data transmitted from vehicle B to vehicle A during the test. However, the following data is used for debugging purposes only. It is hidden from the driver. The 1st column shows the date, 2nd column stores the time taken in 24 hours format, 3rd column is the unique ID from vehicle B which is "Client 1". The 4th column is the time difference in terms of speed in kmh. Negative value means the vehicle B is moving faster than vehicle A. The 5th and 6th column is the difference of longitude and latitude value between vehicle B and vehicle A. The negative value of the longitude shows the vehicle B is moving towards the right side. The last column, latitude with a positive value indicates that the vehicle B is front of vehicle A.

TABLE IX. VEHICLE TRACKING SYSTEM

```
05-04-12 , Client1 , 11:35:29 , -5.092995 , 0.00017 , -0.00006
05-04-12 , Client1 , 11:35:29 , -5.092995 , 0.00017 , -0.00006
05-04-12 , Client1 , 11:35:30 , -3.223 , 0.00019 , -0.00006
05-04-12 , Client1 , 11:35:30 , -3.223 , 0.00019 , -0.00006
05-04-12 , Client1 , 11:35:31 , -0.5550003 , 0.00020 , -0.00006
05-04-12 , Client1 , 11:35:31 , -0.5550003 , 0.00020 , -0.00006
```

Table X and Table XI are the RSSI value on a field test within line of sight. The error rate for Wi Fi Packet Error rate is within 10%. A 5dBi and 2dBi omnidirectional antenna is used for Wi Fi and ZigBee® devices. The test results have shown that the trilateration range estimation for ZigBee® is within ±3 meters tolerance.

TABLE X. WIFI RSSI*

Wi Fi (meters)	Transfer rate	RSSI (dB) - OpenWRT Router with 5dBi omnidirectional antenna	Packet Error Rate(PER)
10	≤54Mbps	-65dBm	<10%
50	≤48Mbps	-67dBm	
80	≤36Mbps	-70dBm	
100	≤12Mbps	-80dBm	

*Maximum Wi Fi client to router, Wi Fi router to router enumeration = 5 second

TABLE XI. ZIGBEE® RSSI*

ZigBee® (meters)	RSSI (dB) – with 2dBi omnidirectional Antenna	Trilateration error Tolerance
10	-40dBm	±3 meters
50	-58dBm	
80	-72dBm	
100	-83dBm	

*Maximum ZigBee® Node to node enumeration = 30ms

Fig. 6 and 7 shows the data transfer rate against distance for Wi Fi and RSSI against distance for ZigBee® signal respectively. The experimental results were obtained within line of sight of two vehicles on an express way during the test drive. The Wi Fi transfer rate from *Vehicle 1* Wi Fi client to *Vehicle 2* Wi Fi router is directly proportional to the distance. As for the ZigBee® router on *vehicle 1* and *2*, the data transfer rate remains consistent at 250Kbps as wireless signal strength directly proportional to the distance.

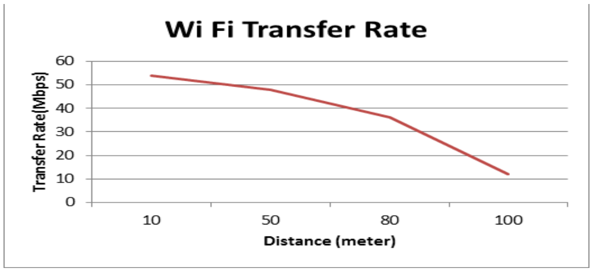


Figure 6. Wi Fi distance (meter) against transfer rate (Mbps)

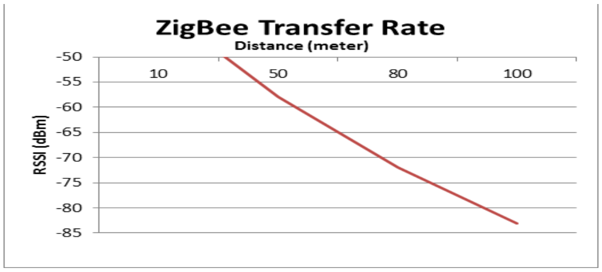


Figure 7. ZigBee® Distance (meter) against RSSI (dBm)

Table XII shows the results for the time taken for vehicle A to search for vehicle B within 100 meter radius. The total time taken to detect the vehicle in normal road condition has been reduced tremendously from 4.6 sec to 426 millisecond with the used of parallel CPU and GPU processing. In term of speed, it has significantly improved approximately 10 times on the system performance.

D. *On road conditions*

1) *Scenario 1 Express way and road tunnel condition*

Under normal traffic conditions, vehicles are communicating in a star topology protocol. There are two different wireless connections each vehicle, Wi Fi and ZigBee® signal.

TABLE XII. PARALLEL CPU AND GPU ACCELERATION

Condition	Without Parallel CPU and GPU(millisecond)	With Parallel CPU and GPU (millisecond)
Cornering	≤4645	≤426
Tail gating	≤1787	≤289
Break Down at middle of road	≤1812	≤335
Lane Change	≤1884	≤383
Blind Spot	≤1921	≤399
Speed Track	≤2779	≤438

While waiting for the Wi Fi client to link up with other vehicles, ZigBee® coordinator and ZigBee® router have established the connection and estimated the trilateration distance between vehicle to vehicle and transmit the location through ZigBee® from vehicle to vehicle. At the same time, GPS location is transmitted through Wi Fi signal. If the vehicle is in a tunnel, the GPS will not be functional. Thus, the distance estimation will rely solely on ZigBee® RSSI from one vehicle to another.

Fig. 8 shows how the communication of Vehicle 1 to 5 is established on normal express way. Each vehicle is linked up to 30 vehicles either on a single hop Wi Fi router to Wi Fi client communication or Wi Fi router to Wi Fi router multi hop communication.

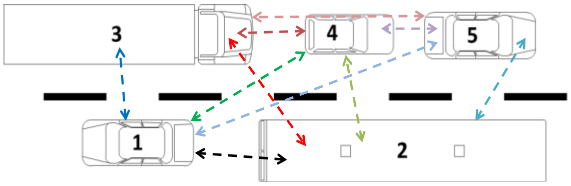


Figure 8. V2V - Normal condition

Table XIII shows how V2V communicate with each other on a road. From Fig. 8, Vehicle ID=1 communicates with the other vehicle ID=2, 3, 4 and 5, vehicle ID=2 communicates with vehicle ID=1, 3, 4 and 5 in a star topology. Similarly for vehicle ID=3, Vehicle ID = 4 and Vehicle ID = 5.

TABLE XIII. V2V COMMUNICATION

Vehicle ID	Vehicle Communication
1	2,3,4,5
2	1,3,4,5
3	1,2,4,5
4	1,2,3,5
5	1,2,3,4

2) *Scenario 2 Traffic jam condition (ZigBee® off)*

In this scenario, the system will pin point the vehicle current location using GPS solely. The GPS location from one vehicle to another is transmitted through Wi Fi until it reaches 100m radius or maximum of 30 vehicles (whichever comes first) on a single hop transmission. Although the Wi Fi takes a few seconds to establish a connection with the client, it is still reasonable for the driver to wait for the time delay in a traffic congested area.

3) *Scenario 3 Indoor car park condition (GPS off)*

The 3rd scenario is in an indoor car park as shown in Fig. 9. ZigBee® routers and ZigBee® coordinators from vehicle 2 to 7 remained functioning even though the vehicle engine is switch off; the RSSI data is transmitted on every second interval. The car park is normally divided into different sections. Assume there are 10 vehicles parked on both left and right row. As vehicle 1 moved towards the direction of vehicle 2 and 5, the RSSI from vehicle 2 to 7 is received by the ZigBee® router of vehicle 1. The estimated error is ±3 meters, which is approximately 2 vehicle space.

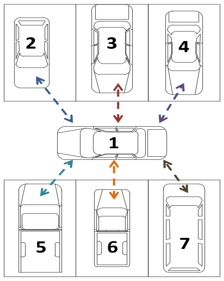


Figure 9. V2V - Indoor car park

VI. CONCLUSIONS

This paper has presented an efficient V2V communication implementation for driver safety by implementing the following methods: 1) Using a variety of real time V2V wireless communication protocols in different circumstances that optimizes performance, and 2) utilizing parallel CPU and GPU processing to accommodate real time processing of massive amounts of data. The methods implemented here are an improvement on past solutions due to the lower cost and lower power consumption of the proposed system without sacrificing performance.

The results in Table XII have shown that, by using parallel CPU and GPU, the time taken can be decreased up to 10 times of normal CPU alone implementations without parallel processing. However the results can further be improved when better hardware CPU and GPU options (in terms of performance per cost) become available. In comparison to other related work in the domain our results in Table XIV shows that the techniques implemented here are lower costs using lower power but with higher or similar speed performance.

Looking forward, the next objective is to port the current system to a mobile device such as a multi core OS hand phone. This porting will further reduce the cost and at the same no additional processing hardware is required other than the mobile phone. The total cost can further be reduced and the design can have much lower power consumption in order to achieve the same performance as the system discussed here.

TABLE XIV. COST AND PERFORMANCE COMPARISON

Criteria	**Results for this research**	**GPS, GPRS/GSM system [16] [17] [18] [19] et. al**	**Radar/Ladar, camera, radio based system [20] [21] [22] [23] et. al**	**WAVE/RSU system [24] [25] [26] [27] et. al**
Total cost	≤USD950	≤USD400	≤USD10K	≤ USD14K
Power consumption	≤150W[a] ≤ 1W for ZigBee® transceiver	≤50W[a]	≤150W[a]	Depending on RSU [44]
Accuracy	±3 meters	±15 meters	±1 meter	±1 meter
Speed	≤ 0.5 second	≤5 second	≤ 50 millisecond Radar [43] response time and ≤ 2 second total time[b]	≤ 2 second[b]

a. Whole System running at full load

b. Depending on the system configuration

REFERENCES

[1] UN, "Road Safety Report," 2012. [Online]. Available: http://www.un.org/ar/roadsafety/pdf/roadsafetyreport.pdf.

[2] C. F. W. X. Yuliya Kopylova, "Accurate Accident Reconstruction in VANET," in Data and Applications Security and Privacy XXV , Springer Berlin Heidelberg, 2011, pp. 271-279.

[3] C. Z. D. T. Soyoung Park, "ReliableTraffic Information Propagation in Vehicular Ad hoc Networks," in Reliable Traffic Popagation, vol. 197, World Scientific , 2008, pp. 1-29.

[4] "Driving Hazards Message Propagation Using Road Side Infrastructure in VANETs," in Advances in Computer Science and Information Technology. Computer Science and Information Technology , Springer Berlin Heidelberg, 2012, pp. 1-8.

[5] S. M. M. Zaydoun Y Rawashdeh, "A novel algorithm to form stable clusters in vehicular ad hoc networks on highways," EURASIP Journal on Wireless Communications and Networking, pp. 1-13, 2012.

[6] T. D. L. Anna Maria Vegni, "Hybrid vehicular communications based on V2V-V2I protocol switching," International Journal Vehicle Information and Communication Systems, vol. 2, no. 3, pp. 213-231, 2011.

[7] M. F. A. H. K. K. H. S.-W. Y. Ali Tufail, "An Empirical Study to Analyze the Feasibility of WIFI for VANETs," in IEEE press, 2008.

[8] S. P. Vaishali D. Khairnar, "V2V communication survey - (Wireless technology)," International Journal of Computer Technology & Applications, vol. 3, no. 1, pp. 370-373, 2012.

[9] S. J. KLAMPÁR M., "Design of communication systems among vehicles," in Proceedings of the 18th Conference EEICT, 2012.

[10] J. G. J. P. B. V. C. G. E. O. J. A. V. Milan´es, "V2I-Based Architecture for Information Exchange among Vehicles," in IFAC Symposium on Intelligent Autonomous Vehicles, 2010.

[11] H. Hartenstein, in VANET - Vehicular Applications and Inter-Networking Technologies, JohnWiley & Sons Ltd., 2010, p. 4.

[12] J. A. E. Toutouh, "Performance analysis of optimized VANET protocols in real world tests," in Wireless Communications and Mobile Computing Conference (IWCMC), 4-8 July 2011 .

[13] A. A. M. S. M. S. k. N. G. N. Q. a. A. S. Muhammad Asif khan, "A Survey on Architecture, Protocols, Challenges and," Internationl journal of multidisciplinary sciences and engineering, vol. 3, no. 3, pp. 5-8, 2012.

[14] I. A. Z. S. K. M. S. M. S. A. J. R. M. A. Rahim, "A comparative study of mobile and vehicular Ad-Hoc Networks," International Journal of Recent Trends in Engineering, vol. 2, no. 4, pp. 195-197, 2009.

[15] B. Mughal, A. Wagan and H. Hasbullah, "Efficient congestion control in VANET for safety messaging," in Information Technology (ITSim), 2010 International Symposium, 2010.

[16] F. I. M. C. B. G. P. D. G. K. Giorgio Rusconi, "I-WAY, intelligent co-operative system for road safety," in IEEE Intelligent Vehicles Symposium, Istanbul, Turkey, 2007.

[17] M. A. Al-Khedher, "Hybrid GPS-GSM Localization of Automobile Tracking System," International Journal of Computer Science & Information Technology, vol. 3, no. 6, pp. 75-85, 2011.

[18] R. K. Neha Verma, "Efficient Data Delivery For Secured Communication in Vanet," Journal of Computer Engineering, vol. 2, no. 2, pp. 1-8, 2012.

[19] D. J. P. H. L. S. L. Z. Yong Li, "Revealing Contact Interval Patterns in Large Scale Urban vehicular ad hoc networks," in Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication , New York, NY, USA, 2012.

[20] K. T. Y. K. S. M. K. K. H. Park S. J., "A novel signal processing technique for vehicle detection radar," in Proceedings on IEEE MTT-S International Microwave Symposium Digest, 2003.

[21] A. M. E.Coelingh, "Collision warning with full autobrake and pedestrian detection — a practical example of automatic emergency braking," in Proceedings of the 13th Internationa lIntelligent Transportation Systems (ITSC) IEEE Conference, 2010.

[22] M. L. X. Q. F. L. Q. L. Tingting Zhou, "Trajectory GenerationModel-Based IMMTracking for Safe Driving in Intersection Scenario," International Journal of Vehicular Technology, pp. 1-8, 2011.

[23] C. W. G. Caicai, "Ground moving target tracking with VS-IMM particle filter based on road information," in Proceedings of the IET International Radar Conference, 2011.

[24] O. K. H. H. Thomas Mangel, "5.9 GHz inter-vehicle communication at intersections: a validated non-line-of-sight pathloss and fading model," Journal on Wireless Communications and Networking, pp. 1-11, 2011.

[25] H. A. C. A. M. R. S. Claudia Campolo, "Impact of Urban Radio Obstructions on effectiveness of moving WAVE providers," in Wireless Communications and Mobile Computing Conference (IWCMC), Istanbul, 2011.

[26] A. M. Claudia Campolo, "Improving V2R connectivity to provides ITS applications in IEEE 802.11p/ WAVE VANET'S," in Telecommunications (ICT), 2011 18th International Conference, Ayia Napa, 2011.

[27] M. J. Annette Bohm, "Real-Time Communication Support for Cooperative, Infrastructure-Based Traffic Safety Applications," International Journal of Vehicular Technology, pp. 1-17, 2011.

[28] "CUDA," NVidia , [Online]. Available: http://www.nvidia.com/object/cuda_home_new.html.

[29] v. studio, "www.microsoft.com/visualstudio/," [Online]. Available: www.microsoft.com/visualstudio/.

[30] WCF. [Online]. Available: http://wcf.codeplex.com/.

[31] ".Net Framework," Microsoft, 2011. [Online]. Available: http://www.microsoft.com/net/.

[32] WMI. [Online]. Available: http://msdn.microsoft.com/en-us/library/windows/desktop/aa394582(v=vs.85).aspx.

[33] "Windows SDK for Windows 7," Microsoft, 2011. [Online]. Available: http://www.microsoft.com/download/en/details.aspx?id=8279.

[34] "OpenWRT," 2011. [Online]. Available: https://openwrt.org/.

[35] "ZigBee," ZigBee Alliance, 2012. [Online]. Available: http://www.zigbee.org/.

[36] "Ember," ZigBee, 2012. [Online]. Available: http://www.ember.com/products_zigbee_software.html.

[37] "GPS 18x," Garmin, [Online]. Available: http://www8.garmin.com/manuals/GPS18x_TechnicalSpecifications.pdf NHBfbZm5ZXtJ13QNpbnjGCMnP7GEQ.

[38] "Broadband modem," HuaWei, 2012. [Online]. Available: http://www.huawei.com/.

[39] "Threads Building blocks," Intel, 2011. [Online]. Available: http://threadingbuildingblocks.org/.

[40] H. S. a. A. C. BhavneetSidhu, "Emerging Wireless Standards - WiFi, ZigBee," in World Academy of Science, Engineering and Technology, 2007.

[41] NVidia, "CUDA Memories," [Online]. Available: http://courses.engr.illinois.edu/ece498/al/textbook/Chapter4-CudaMemoryModel.pdf.

[42] NVidia, "NVidia CUDA C Programming Guide," 2012. [Online]. Available: http://docs.nvidia.com/cuda/index.html.

[43] "ESR," 2012. [Online]. Available: http://www.autonomoustuff.com/uploads/9/6/0/5/9605198/radar_comparison_chart_as.pdf.

[44] "Volpe," 2012. [Online]. Available: http://www.volpe.dot.gov/coi/att/work/safety-pilot.html.

Simple Method for Ontology Automatic Extraction from Documents

Andreia Dal Ponte Novelli
Dept. of Computer Science
Aeronautic Technological Institute
Dept. of Informatics
Federal Institute of Sao Paulo
Sao Paulo – Brazil

José Maria Parente de Oliveira
Dept. of Computer Science
Aeronautic Technological Institute
Sao Paulo - Brazil

Abstract—**There are many situations where it is needed to represent and analyze the concepts that describe a document or a collection of documents. One of such situations is the information retrieval, which is becoming more complex by the growing number and variety of document types. One way to represent the concepts is through a formal structure using ontologies. Thus, this article presents a fast and simple method for automatic extraction of ontologies from documents or from a collections of documents that is independent of the document type and uses the junction of several theories and techniques, such as latent semantic for the extraction of initial concepts, Wordnet and similarity to obtain the correlation between the concepts.**

Keywords-document ontology; ontology creation; ontology extraction; concept representation.

I. Introduction

The continuous increase in the amount of documents produced both on the Web and in local repositories makes it increasingly complex and costly to analyze, categorize and retrieve documents without considering the semantics of the whole or each document. Generally, the semantics is analyzed based on the concepts contained in the documents, and the ontologies are one of the ways to represent these concepts. Ontology can be defined as a formal and explicit specification of shared conceptualization [1]. These can also be seen as conceptual models that capture and explain the used vocabulary in semantic specifications. For documents, ontology may be seen as significant group of terms that expresses the vocabulary of the document through concepts and relations modeled after those terms.

Ontology can be constructed in a more general way or a domain-dependent, depending on how general are the sets of concept. In the context of this article, the concepts of ontologies are general, since the terms used for formation of the concept may be present in any document. However, the ontology creation is based on documents from a specific area, thereby resulting ontologies directed to the document domain.

There are many situations where presence of semantics is necessary in order to best perform certain tasks in certain areas, however, depending on the task, it is not necessary that the semantics be extremely detailed regarding the formation of concepts and semantic relations, since the semantics is an auxiliary item to the task. Thus, the proposed method tries to meet the need of creating a simple and meaningful semantic description of documents without analyzing these documents through artificial intelligence techniques, language and context analysis.

The method extracts an ontology from a collection of any documents (text only or structured) or descriptive ontologies of single documents using tools and techniques such as latent semantic analysis, clustering and Wordnet. The initial concepts of the ontology and its relations are obtained from the terms of the documents and other concepts are created from the analysis of the terms using latent semantic and clustering. The relations between the concepts are obtained from analysis using a thesaurus or ontology, and for this work Wordnet was chosen to.

This article is organized as follows: section II presents the state of the art for the automatic extraction of ontologies, in Section III it is presented the concepts of latent semantic analysis, clustering and Wordnet used in this work; it is presented in Section IV the proposed method detailing its operation and experiments, and in section V the conclusions of the article.

II. Related Work

There are many works in the literature that deal with generation or extraction of ontologies. Most of the works focus on certain documents types or on specific domains.

Initially, it is presented solutions related to ontologies generation using algorithms such as clustering and latent semantic that are relatively independent of the document type, since they only use the textual content of the documents for the ontologies creation.

The work of Maddi et al. [2] presents a way to extract ontologies for text documents using singular value decomposition (SVD) to obtain the concepts from terms and represents the obtained results using bipartite graphs.

Fortuna et al. [3] present a process for obtaining concepts semi-automatically, because the solution only suggests terms sets and from this suggestion the user chooses the concepts and makes connections between them.

Still considering the use of latent semantic, Yeh and Yang [4] generate ontologies from historical documents from digital libraries, using latent semantics for generating the initial concepts and clustering for the other concepts. Regarding the semantic relation generation, the paper proposes the use of a specific set of pre-defined relations to the language and document domain.

Some paper presents detailed studies on the generation of concepts and ontologies. Thus, the state of art for methods, techniques and tools to the ontologies generation is presented in [5, 6, 7], and in [3] it is presents a study of concepts generation focused on clustering and latent semantic.

Considering the solutions that generate ontologies for applications and specific document types, there is the work of Sanchez and Moreno [8] that presents a methodology for automatic construction of domain ontologies in which concepts are obtained of keywords from Web pages. The ontologies creation for lecture notes in distance education systems is presented in [9] and it uses natural language processing to extract keywords, algorithms based on frequency to select the concepts from the keywords and association rules algorithms to define the semantic relations.

Gillam and Ahmad [10] propose the obtainment of concepts using statistical methods for comparison between a vocabulary created by domain experts and the general vocabulary words from the text. For the hierarchy creation it is used solutions from literature, such as smoothing and extraction and placement technique.

Lee et al. [11] present a solution for creating ontologies from text document in Chinese using fuzzy logic, similarity and clustering to obtain the taxonomy of the ontology.

The works presented in the literature are generally directed to a particular area or document type, whereas the proposed method is developed to meet different domains and document types.

Most solutions in the literature generate, as an answer, an ontology that can be manipulated by only using the tool that develops the solution, limiting the use of ontology developed or requiring an adaptation for use in other environments. Thus, the proposed method generates a standard OWL ontology that can be accessed and manipulated in ontology editors or other tools, for example, Gena when programmed in Java.

Another consideration that must be made about the solutions for creating ontologies is that solutions from the literature require the intervention of a specialist to obtain the semantic relations or algorithm that take much time and effort. Therefore, the proposed method uses a simple and relatively quick way to automatically generate the basic semantic relations between the concepts, generating an ontology that has the properties, axioms and constraints on its outcome.

III. CONCEPTS

In this section, it is presented some concepts and techniques used in the development of the proposed method.

A. Latent Semantic Analysis and Singular Value Decomposition (SVD)

Latent Semantic Analysis is a way to manipulate sets of documents [12]. However, in the context of this work, it is used to obtain concepts that comprise a set of documents [2].

The latent semantic analysis explores the relation between terms and documents to build a vector space, which allows the performing of analyzes between documents. To apply the latent semantic index-terms must be obtained which are the most frequent terms in the documents. From the index terms, it is mounted a term-document matrix containing the terms in rows and the term frequency in columns for each of the documents. As the document-term matrix can be very large to be fully analyzed, the SVD is used to obtain an approximation of this matrix through linear combinations.

The SVD decomposes the term-document matrix into three matrices U, Σ and V, where U is an orthonormal matrix whose columns are called singular vectors to the left, Σ is a diagonal matrix whose elements are called not negative singular values and V is an orthonormal matrix whose columns are called singular vectors to the right. Fig. 1 shows the decomposition of an A document-term matrix with dimensions mxn, resulting in matrices U with dimensions mxr, Σ with dimensions rxr and V with dimensions rxn.

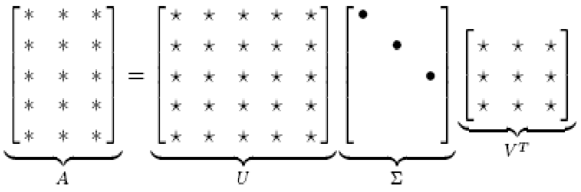


Figure 1. Example of singular value decomposition for a term-document matrix A.

The use of SVD allows both dimensionality reduction of term-document matrix for an information recovery task and the creation of concepts and their association with the document.

The creation of concepts is performed by analyzing the two matrices term-document and U. The first level (ground level) of the ontology hierarchy is obtained from its own index terms from the term-document matrix. The next level of the hierarchy is formed of concepts obtained from the term analysis of the matrix U columns, which provides the relation between terms and concepts. A concept consists of chosen terms from each column according to some criterion.

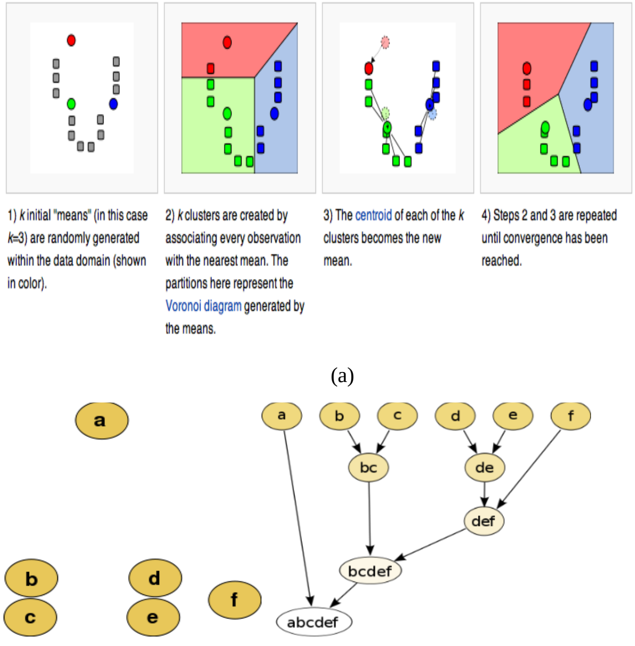
The matrix V provides the relation between concepts obtained from the U matrix and the documents of collection, allowing one to know which concepts are from each document and create a descriptive ontology for each document.

B. Hierarchical Clustering Algorithms

The clustering algorithms in the context of this method are used to perform an analysis on concepts obtained to generate the other levels of the hierarchy of the ontologies. Thus, this section presents the concepts related to hierarchical clustering.

There are two ways to implement hierarchical clustering: bottom-up and top-down. The bottom-up solution starts with several individual concepts that are grouped together with more similar ones until it forms a single group. On the other hand, a top-down solution starts with all objects in one group and these are subdivided according to their proximity in smaller groups.

Among the various bottom-up clustering algorithms, there are two that are most commonly used for creating ontology hierarchies. The K-Means algorithm was presented in 1967 and it begins at the choice of baseline groups (centroids). The algorithm works by arranging objects according to these centroids and recalculating these centers until the result of convergence is satisfactory [13]. However, the clustering algorithm initially considers that all objects are separate groups. The algorithm analyzes the similarity between the two groups putting them together based on the proximity between the groups until there is only one group. Fig. 2 illustrates the operation of K-Means clustering algorithms and clustering.



1) *k* initial "means" (in this case *k*=3) are randomly generated within the data domain (shown in color).

2) *k* clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.

3) The centroid of each of the *k* clusters becomes the new mean.

4) Steps 2 and 3 are repeated until convergence has been reached.

(a)

(b)

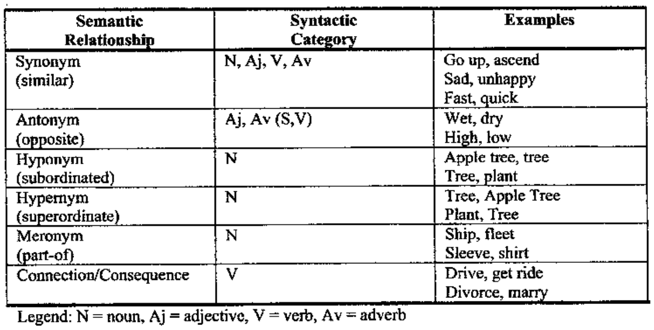
Figure 2. Example of operation of clustering algorithms that can be used in the building of ontologies [13].

C. Wordnet Ontology

Wordnet [14] may be considered an ontology constructed in a more general way or also a lexical reference which can be used online or locally. According to Snasel et al. [15], Wordnet has information of nouns, adjectives, verbs and adverbs, which can be used to determine semantic connections and to trace the connections between morphological words.

Generally, there is a version of Wordnet for each language. However, there are tools to extend the analysis in one language to others, for example, if the noun "house" is analyzed to obtain synonyms, using the tool, all its synonyms may be obtained for English or for any other language.

In this method context, Wordnet is used to create the semantic relations between the ontology concepts focusing on the creation of properties, axioms and restrictions. For the creation of these relations are analyzed possible relations proposed in Wordnet, as shown in Fig. 3.



Semantic Relationship	Syntactic Category	Examples
Synonym (similar)	N, Aj, V, Av	Go up, ascend Sad, unhappy Fast, quick
Antonym (opposite)	Aj, Av (S,V)	Wet, dry High, low
Hyponym (subordinated)	N	Apple tree, tree Tree, plant
Hypernym (superordinate)	N	Tree, Apple Tree Plant, Tree
Meronym (part-of)	N	Ship, fleet Sleeve, shirt
Connection/Consequence	V	Drive, get ride Divorce, marry

Legend: N = noun, Aj = adjective, V = verb, Av = adverb

Figure 3. Wordnet Semantic Relationships [16].

IV. Proposed Method

The proposed method presents a simple, rapid and automatic way of obtaining an initial organization of concepts from collection of any documents that can be formed only by text or structure and text. This proposal aims to meet applications that require semantic descriptions that are meaningful only enough to meet the application and does not need much detail. This method improves some solutions that make use of clustering and statistical methods in order to obtain more significant ontologies by improving the development of concepts and semantic relations. In this method it is possible to obtain an ontology that describes the concepts of an individual document or of a collection of documents. The method seeks to work only on an automated way, making a specialist unnecessary at the time of the ontology creation. However, a domain expert may do an analysis of the ontology using an ontology editor and make changes to improve the result obtained automatically. The method also keeps stored summaries and elements used to obtain the concepts and terms, so that this method allows the inclusion of new documents in the collection, as well as the deletion and alteration.

Fig. 4 shows the method general outline of ontologies extraction from a collection of documents. In the following sections the main parts of the method and the results obtained using it are shown.

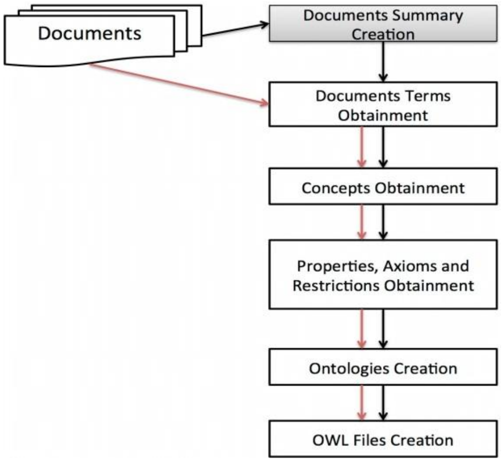

Documents
Documents Summary Creation
Documents Terms Obtainment
Concepts Obtainment
Properties, Axioms and Restrictions Obtainment
Ontologies Creation
OWL Files Creation

Figure 4.General outline of the proposed method operation.

A. Documents Preparation

In this first phase, documents are prepared for obtaining concepts. First, it is necessary that the collection of documents be analyzed in order to define which documents have structure (XML documents) or text only.

Considering only text documents, initially, it is analyzed the necessity of obtaining a summary if the document is very large. For summaries preparation can be used one of the algorithms present in the literature depending on the desired quality and efficiency. The summaries of documents can be kept stored in files or databases, since their preparation need a reasonable computational time that can be suppressed by keeping them for use in the preparation of other ontologies when these documents are used.

For documents that have structure, there is a prior step to the summaries creation. This step is the separation between structure and content of the document. In this separation, the structure is analyzed to verify if the elements have definitions that can be considered concepts in the ontology. The elements are ignored if the structure does not have relevant ones, otherwise they are also stored. The separated contents are analyzed following the same idea of only text documents.

The summaries / documents are read, extracting the terms that will be used in the preparation of the ontology, i.e., these are transformed into set of strings containing terms not repeated and considered relevant of each of the summaries / documents.

These terms also undergo a standardization process, that is, the terms are analyzed in order to withdraw from the set terms that are grammatically different forms for the same word, such as student and students, and terms that are different tenses for the same verb, for example, walk and walks. For XML documents, the term set can contain structure elements that are relevant to the formation of concepts.

Still at this stage, the terms need to have their TF-IDF (Term Frequency Inverse Document Frequency) calculated. The TF-IDF is calculated in two steps, first TF is obtained by the formula presented in (1):

TF= freq_(i,j)/max(freq_(l,i)) (1)

where freq_(i,f) is the frequency of term i for a document j and e max(freq_(l,j)) is the frequency of the most frequent term in the document. However, the IDF is the second stage of the calculation, and it is obtained by the formula (2) shown below:

IDF= log 〖N/n_i 〗 (2)

where N is the total number of documents of the set and n_i is the number of documents that contain term i. The final result of TF-IDF is obtained by multiplying the TF by the IDF. The TF-IDF is used in the next phase, in getting the concepts.

B. Concepts Obtainment

Initially, it is obtained the index terms, which are the set of terms that appear in more than twenty five percent of the documents. If this obtained set of terms is very large, it can be reduced by selecting a subset of these terms observing the criterion of keeping in the index the terms that appear more frequently in the documents, so that the manipulation of the document-term matrix and of the matrices created by SVD become easier.

For the resulting matrices from the application of SVD in term-document matrix, the matrices U that links the concepts to the terms and V matrix that links the concepts to documents are used.

The use of the matrix U in order to obtain the ontology concepts has been shown in [2, 3]. The concepts are created from the terms of the matrix U columns. Thus, each column from U creates a concept from the union of the terms that have the highest values in the column, with maximum of three terms united. A comparative analysis is made in this obtained set of concepts in order to verify the concepts that may be the same, i.e., those having the same terms only placed differently. If the concepts are actually different, they are kept in set of concepts and the terms are attached to these concepts. The linking between the terms and concepts is done through analysis of matrix U, verifying in each column the terms that have values greater than 0.5, because the relation is only considered valid if the connectivity degree is greater than fifty percent.

The obtained concepts from the matrix U are the ones of the intermediate level of the ontology, that is, the second level in the hierarchy. At the base level of the ontology, there are the initial concepts that are themselves index terms.

Fig. 5 shows an obtained concept in one experiment performed and its terms, with concept being formed of two terms with the greatest value in the matrix U column.

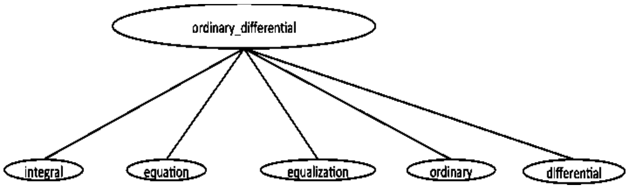
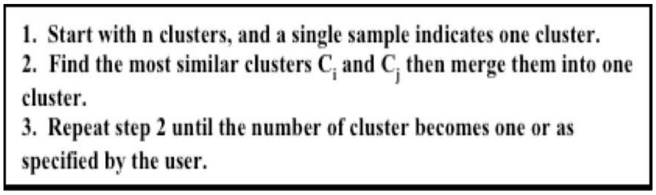


Figure 5.Example of a concept and its terms.

From the two obtained levels of concepts, it is necessary to create the other levels of the ontology to form a complete hierarchy. Thus, it is used the algorithm shown in Fig. 6 for clustering the concepts until obtaining an only group which will be the main concept of hierarchy.



1. Start with n clusters, and a single sample indicates one cluster.
2. Find the most similar clusters C_i and C_j then merge them into one cluster.
3. Repeat step 2 until the number of cluster becomes one or as specified by the user.

Figure 6. Agglomerative Algorithm.

At this point you need to check which concepts belong to which documents, because only the concepts of each one of the documents are clustered. Thus, before clustering, it is necessary that the matrix V be analyzed by separating the concepts and terms of each of the documents. As the terms, the concepts also have its degree of connectivity analyzed, and it is considered for the document only those concepts that have connectivity superior to fifty percent, therefore ensuring greater quality in the developed ontology.

C. Creation of properties, Axioms and Restrictions

After defining the concepts of each document, it is obtained the semantic relations for each one of the ontologies. These relations are organized into properties, axioms and constraints. Two types of properties can be defined: the object properties and data type properties. Object properties relate instances with other instances defining restrictions and behaviors. Data types refer to properties that express only values, e.g., strings or numbers. The concepts can have super and sub-concepts, providing a rationalizing mechanism and property inheritance. Finally, the axioms are used to provide information about the concepts and properties, such as, to specify the equivalence of two concepts or range of a property.

There are many semantic relations that can be obtained using Wordnet. Initially, it is set up the simplest of the properties, which is the subclass_of between concepts of different levels that form the ontology. After, other relations like, equivalent_to (between synonyms or similar concepts), disjoin_of (between antonyms), part_of (between terms that complete others) and inverse_of (between antonyms and synonyms), can be defined. To define these relations, the concepts are analyzed using Wordnet, verifying possible correlations between the considered concepts. For these found correlations, it is analyzed the ones which are suitable for the use in the ontology definition, for example, if the concepts are synonyms, they are given an equivalence defined axiom. In this work, only Wordnet ontology was used to obtaining these correlations, however, depending on the document field, other ontologies may be used.

Besides Wordnet, the concepts are also analyzed for their degree of similarity. Depending on this similarity value, the concepts receive the semantic relation of equivalence. For this work, it was accepted as equivalent the concepts which have a degree of similarity greater than 0.90.

To simplify the process of the semantic relations obtainment, the analysis is performed by level, i.e., the concepts of a same level are examined in pairs until all possible relations are defined.

Fig. 7 presents the semantic relations defined for the concept in Fig. 5, being these relations are: subclass_of between the concept and the terms, and equivalant_to for synonymous terms.

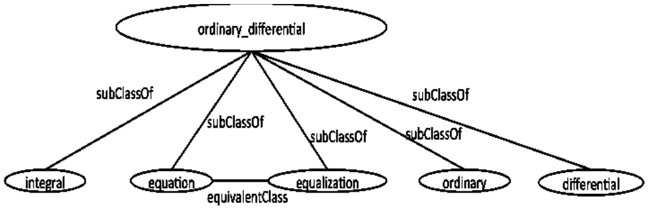


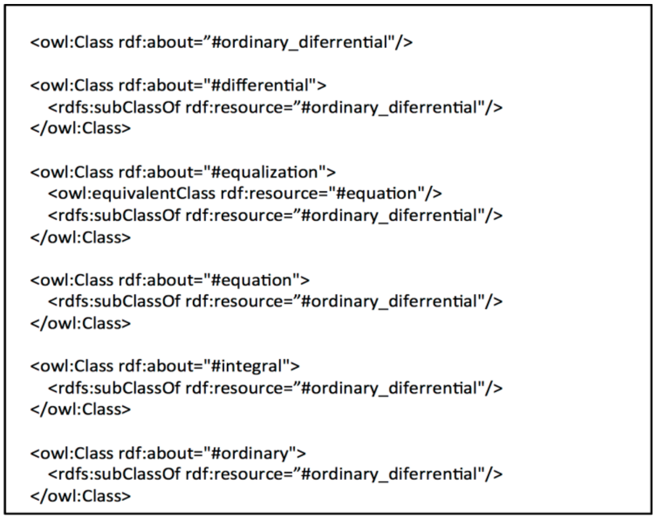
Figure 7. Example of defined properties for a concept.

D. Ontology Creation

This is last phase of method. The concept and semantic relations are organized in ontologies that are stored in files encoded in OWL language. This language is used to define ontologies and it provides mechanisms for component creation: concepts, instances, properties and axioms.

As a result of this phase, there is a set of ontologies in OWL for the documents in the collection. However, using all concepts and relations obtained, a single ontology describing the entire collection can be created, thus creating an ontology that may be worked by a specialist to form an ontology of domain.

Fig. 8 shows an example of a possible OWL coding to the concepts of Fig. 7.

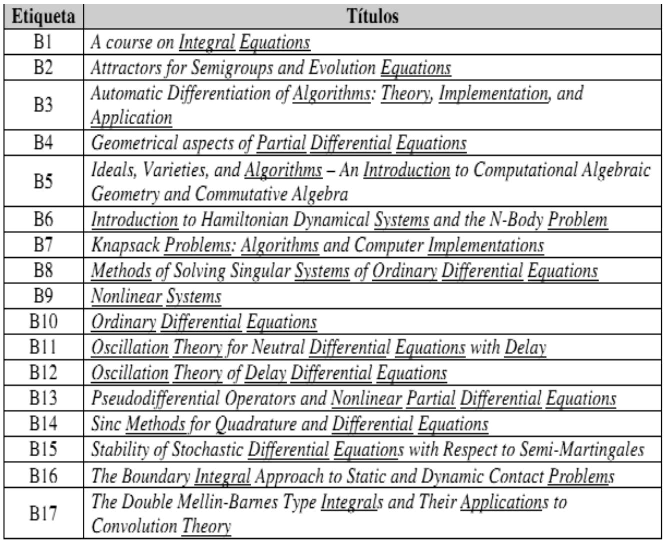


```
<owl:Class rdf:about=”#ordinary_diferrential"/>

<owl:Class rdf:about="#differential">
  <rdfs:subClassOf rdf:resource=”#ordinary_diferrential"/>
</owl:Class>

<owl:Class rdf:about="#equalization">
  <owl:equivalentClass rdf:resource="#equation"/>
  <rdfs:subClassOf rdf:resource=”#ordinary_diferrential"/>
</owl:Class>

<owl:Class rdf:about="#equation">
  <rdfs:subClassOf rdf:resource=”#ordinary_diferrential"/>
</owl:Class>

<owl:Class rdf:about="#integral">
  <rdfs:subClassOf rdf:resource=”#ordinary_diferrential"/>
</owl:Class>

<owl:Class rdf:about="#ordinary">
  <rdfs:subClassOf rdf:resource=”#ordinary_diferrential"/>
</owl:Class>
```

Figure 8. OWL codification to the concepts of Fig. 7.

E. Experimental Results

To validate the proposed method, ontologies were created for both text and XML collections of documents. The descriptions of collections and of obtained results are provided below. The first experiment creates ontologies for a simple collection of documents with small texts about book titles. The group has seventeen documents, as shown in Fig. 9.

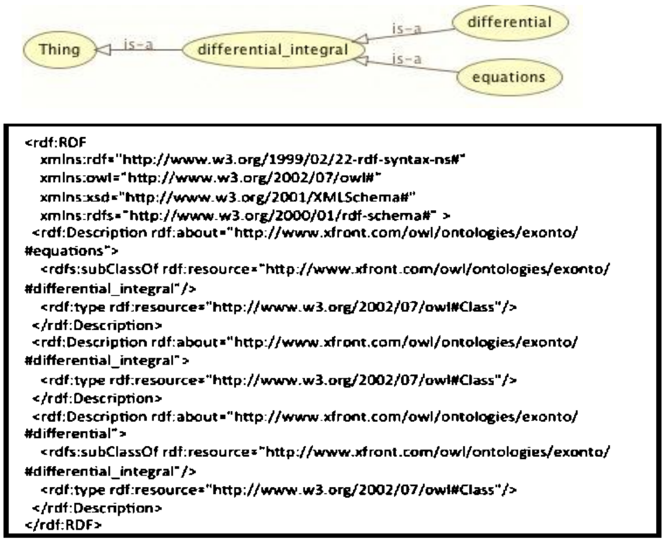


Etiqueta	Títulos
B1	*A course on Integral Equations*
B2	*Attractors for Semigroups and Evolution Equations*
B3	*Automatic Differentiation of Algorithms: Theory, Implementation, and Application*
B4	*Geometrical aspects of Partial Differential Equations*
B5	*Ideals, Varieties, and Algorithms – An Introduction to Computational Algebraic Geometry and Commutative Algebra*
B6	*Introduction to Hamiltonian Dynamical Systems and the N-Body Problem*
B7	*Knapsack Problems: Algorithms and Computer Implementations*
B8	*Methods of Solving Singular Systems of Ordinary Differential Equations*
B9	*Nonlinear Systems*
B10	*Ordinary Differential Equations*
B11	*Oscillation Theory for Neutral Differential Equations with Delay*
B12	*Oscillation Theory of Delay Differential Equations*
B13	*Pseudodifferential Operators and Nonlinear Partial Differential Equations*
B14	*Sinc Methods for Quadrature and Differential Equations*
B15	*Stability of Stochastic Differential Equations with Respect to Semi-Martingales*
B16	*The Boundary Integral Approach to Static and Dynamic Contact Problems*
B17	*The Double Mellin-Barnes Type Integrals and Their Applications to Convolution Theory*

Figure 9. Presentation of the experiment documents and their contents [12].

As it is a collection with very short texts there is no need to create summaries. Thus, the method begins obtaining the terms sets of documents, on which it is applied the latent semantic technique and the other method steps for the building of ontologies.

Fig. 10 shows encoding OWL of ontology of document B4 and its graph generated in the Protégé editor available at [17]. Fig. 11 shows created ontology of document B11 where it can be seen a larger number of semantic relations between elements.



```
<rdf:ROF
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:owl="http://www.w3.org/2002/07/owl#"
   xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" >
  <rdf:Description rdf:about="http://www.xfront.com/owl/ontologies/exonto/
#equations">
   <rdfs:subClassOf rdf:resource="http://www.xfront.com/owl/ontologies/exonto/
#differential_integral"/>
   <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.xfront.com/owl/ontologies/exonto/
#differential_integral">
   <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.xfront.com/owl/ontologies/exonto/
#differential">
   <rdfs:subClassOf rdf:resource="http://www.xfront.com/owl/ontologies/exonto/
#differential_integral"/>
   <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#Class"/>
  </rdf:Description>
</rdf:RDF>
```

Figure 10. Example of generated ontology for document B4 of the collection presented in Fig. 9.

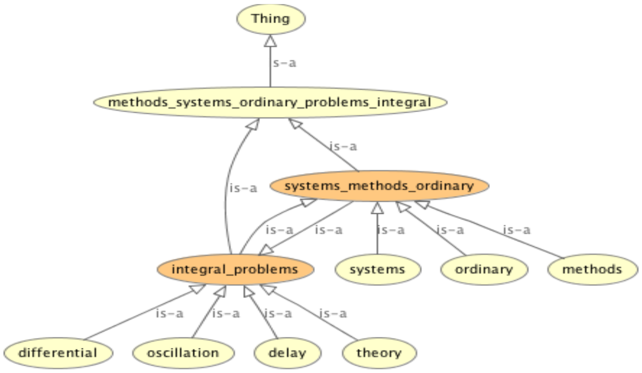
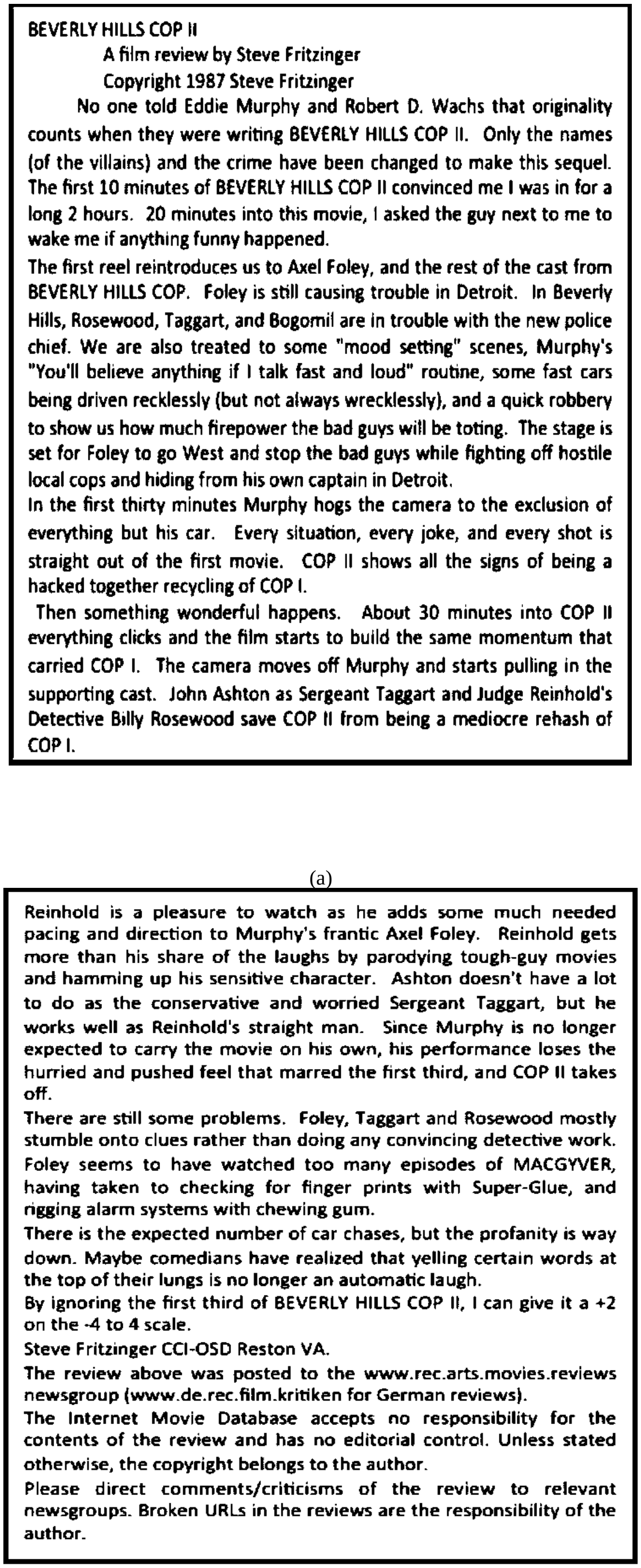


Figure 11. Generated ontology for document B11 of the collection presented in Fig. 9.

Still considering only text documents, a second experiment was carried out using documents with larger size, which require the preparation of a summary. The collection has fifteen documents chosen randomly from a collection of twenty-seven thousand documents about international movie reviews. From this collection, the ontologies for each one of the documents and for the whole collection of documents were obtained. Fig. 12 shows in (a), a text of document example; in (b), terms of the generated summary for the document; and in (c), the ontology created for the document.



BEVERLY HILLS COP II

A film review by Steve Fritzinger

Copyright 1987 Steve Fritzinger

No one told Eddie Murphy and Robert D. Wachs that originality counts when they were writing BEVERLY HILLS COP II. Only the names (of the villains) and the crime have been changed to make this sequel. The first 10 minutes of BEVERLY HILLS COP II convinced me I was in for a long 2 hours. 20 minutes into this movie, I asked the guy next to me to wake me if anything funny happened.

The first reel reintroduces us to Axel Foley, and the rest of the cast from BEVERLY HILLS COP. Foley is still causing trouble in Detroit. In Beverly Hills, Rosewood, Taggart, and Bogomil are in trouble with the new police chief. We are also treated to some "mood setting" scenes, Murphy's "You'll believe anything if I talk fast and loud" routine, some fast cars being driven recklessly (but not always wrecklessly), and a quick robbery to show us how much firepower the bad guys will be toting. The stage is set for Foley to go West and stop the bad guys while fighting off hostile local cops and hiding from his own captain in Detroit.

In the first thirty minutes Murphy hogs the camera to the exclusion of everything but his car. Every situation, every joke, and every shot is straight out of the first movie. COP II shows all the signs of being a hacked together recycling of COP I.

Then something wonderful happens. About 30 minutes into COP II everything clicks and the film starts to build the same momentum that carried COP I. The camera moves off Murphy and starts pulling in the supporting cast. John Ashton as Sergeant Taggart and Judge Reinhold's Detective Billy Rosewood save COP II from being a mediocre rehash of COP I.

(a)

Reinhold is a pleasure to watch as he adds some much needed pacing and direction to Murphy's frantic Axel Foley. Reinhold gets more than his share of the laughs by parodying tough-guy movies and hamming up his sensitive character. Ashton doesn't have a lot to do as the conservative and worried Sergeant Taggart, but he works well as Reinhold's straight man. Since Murphy is no longer expected to carry the movie on his own, his performance loses the hurried and pushed feel that marred the first third, and COP II takes off.

There are still some problems. Foley, Taggart and Rosewood mostly stumble onto clues rather than doing any convincing detective work. Foley seems to have watched too many episodes of MACGYVER, having taken to checking for finger prints with Super-Glue, and rigging alarm systems with chewing gum.

There is the expected number of car chases, but the profanity is way down. Maybe comedians have realized that yelling certain words at the top of their lungs is no longer an automatic laugh.

By ignoring the first third of BEVERLY HILLS COP II, I can give it a +2 on the -4 to 4 scale.

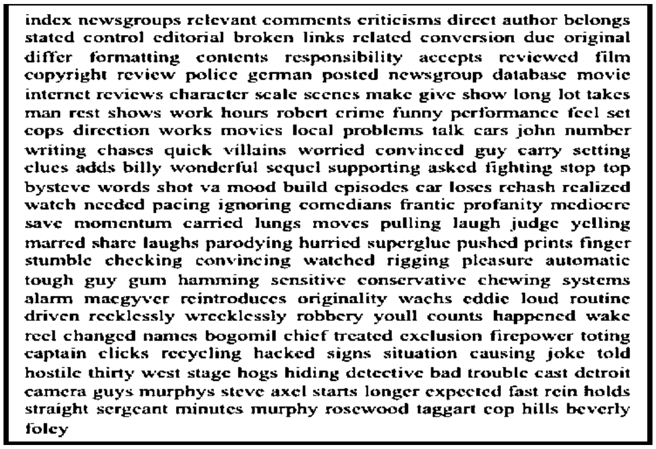
Steve Fritzinger CCI-OSD Reston VA.

The review above was posted to the www.rec.arts.movies.reviews newsgroup (www.de.rec.film.kritiken for German reviews).

The Internet Movie Database accepts no responsibility for the contents of the review and has no editorial control. Unless stated otherwise, the copyright belongs to the author.

Please direct comments/criticisms of the review to relevant newsgroups. Broken URLs in the reviews are the responsibility of the author.

(b)



index newsgroups relevant comments criticisms direct author belongs stated control editorial broken links related conversion due original differ formatting contents responsibility accepts reviewed film copyright review police german posted newsgroup database movie internet reviews character scale scenes make give show long lot takes man rest shows work hours robert crime funny performance feel set cops direction works movies local problems talk cars john number writing chases quick villains worried convinced guy carry setting clues adds billy wonderful sequel supporting asked fighting stop top bysteve words shot va mood build episodes car loses rehash realized watch needed pacing ignoring comedians frantic profanity mediocre save momentum carried lungs moves pulling laugh judge yelling marred share laughs parodying hurried superglue pushed prints finger stumble checking convincing watched rigging pleasure automatic tough guy gum hamming sensitive conservative chewing systems alarm macgyver reintroduces originality wachs eddie loud routine driven recklessly wrecklessly robbery youll counts happened wake reel changed names bogomil chief treated exclusion firepower toting captain clicks recycling hacked signs situation causing joke told hostile thirty west stage hogs hiding detective bad trouble cast detroit camera guys murphys steve axel starts longer expected fast rein holds straight sergeant minutes murphy rosewood taggart cop hills beverly foley

(c)

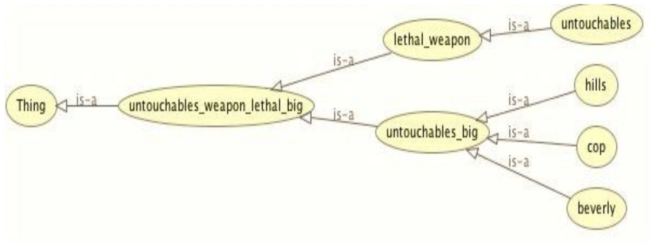


Figure 12. Text example, summary and ontology created. In part (a), the document was presented; in part (b), the summary generated for the document; and in part (c), the ontology for this document.

As presented in the previous sections, it is possible to generate an ontology that describes the concepts and terms of the whole collection of documents. Thus, Fig. 13 shows the generated ontology for the experiment fifteen documents.

The third experiment was carried out with a collection of twenty-four XML documents about historical manuscripts. For the development of this experiment, first it was carried out the separation between structure and content.

After the separation between the documents structure and content, the analysis was performed to verify if the structure had relevant information for the ontology formation. In the case of used documents in experiment, the structure is nothing but the structuring of text sections, so only the content was used.

Since each document in the set has a considerable number of pages, the number of terms in summaries is high, and also the index-terms set, complicating the matrix manipulation at the concept obtainment time. Thus, it was considered for this experiment only the five hundred more frequent terms in the documents to obtain the index-terms. Applying the proposed method, ontologies have been created for each of the documents and for the collection.

Fig. 14 shows the ontology created to the collection of document, demonstrating concepts, terms and semantic relations

The carried out experiments showed that the individual ontologies generated to documents express significantly, even though simply, the concepts contained therein. For example, for the document shown in Figure 12, it is possible to notice that the film described in the presented review has to do with a local (Beverly Hill), police and violence (lethal and weapon). As for the document B4, it is possible to know that the document is a book about some aspect of differential equations as show the ontology concepts, differential and equations.

These experiments demonstrated that the method satisfactorily obtained the semantic relations between concepts and terms simply and automatically, improving the created ontology, because it can be identified similar terms by synonymy and other relations due to the use of similarity and Wordnet. In order to improve the obtained semantic relations in the created ontologies, it would be possible to use other ontologies or a thesaurus besides Wordnet.

The proposed method has fulfilled its proposition because even though it is very simple, the use of Wordnet combined with the employed techniques have improved the obtained results, allowing better definition of document concepts and the semantic relations that compose the generated ontology.

The resulting ontologies are stored in an OWL file that can be edited or viewed by the usual ontology editors, allowing its easier handling.

V CONCLUSION

This paper presented a method for document or collection of documents ontology extraction using latent semantic, clustering and Wordnet. The proposed method is fully automatic and simple, but with significant results enough to allow the understanding and manipulation of the document concepts without needing advanced techniques, the intervention of an expert, or even the entire understanding of the domain.

The experiments showed that the obtained ontologies satisfactorily represent the concepts of the documents. Despite that, this method can still be improved using other tools and techniques that allow the definition of other semantic relations between the concepts and enhance the concepts obtainment.

REFERENCES

[1] K. Breitman, Web Semantica - A internet do Futuro, vol. 1, Rio de Janeiro, RJ: LTC, 2006, p. 190.

[2] G. R. Maddi, C. S. Velvadapu, S. Strivastava e J. G. d. Lamadrid, “Ontology Extraction from Text Documents by Singular Value Decomposition,” em ADMI 2001, 2001.

[3] B. Fortuna, D. Mladenic e M. Grobelniz, “Semi-automatic Construction of Topic Ontology,” em Lecture Notes in Computer Science, vol. 4289, Springer, 2005, pp. 121-131.

[4] J. Yeh e N. Yang, “Ontology Construction on Latent Topic Extraction in a Digital Library,” em International Conference on Asian Digital Libraries 2008, 2008.

[5] Y. Ding e S. Foo, “Ontology Research and Development Part 1 – A Review of Ontology Generation,” Journal of Information Science, vol. 28, pp. 123-136, 2002.

[6] I. Bedini e B. Nguyen, “Automatic ontology generation: State of the art,” 2007.

[7] A. Zauaq, “A survey of Domain Engineering: Method and Tools,” em Studies in Computational Intelligence, vol. 308, 2010, pp. 103-119.

[8] D. Sanchez e A. Moreno, “Creating ontologies from Web documents,” Recent Advances in Artificial Intelligence Research and Development, vol. 113, pp. 11-18, 2004.

[9] N. Gantayat, “Automated Construction of Domain Ontologies from Lecture Notes,” Bombay, 2011.

[10] L. Gillam e K. Ahmad, “Automatic Ontology Extraction from Unstructured Texts,” em the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE, 2005.

[11] C.S. Lee, Y.-F. Kao, Y.-H. Kou e M.-H. Wang, “Automated Ontology Construction for Unstructured Text Documents,” Data & Knowledge Engineering, 2007.

[12] D. Foronda, “Estudo Exploratório da Indexação Semantica Latente,” PUC-RS, Porto Alegre.

[13] [13] B. P. Nunes, “Classificação automatica de dados semi-estruturados,” Rio de Janeiro, 2009.

[14] “Wordnet,” [Online]. Available: http://wordnet.princeton.edu. [Accessed in 2012].

[15] V. Snasel, P. Maravec e J. Pokorney, “Wordnet Ontology based Model for Web Retirval,” em WIRI '05 Proceedings of the International Workshop on Challenges in Web Information Retrieval and Integration, 2005.

[16] K. Breitman, M. A. Casanova e W. Truszkowski, Semantic Web: Concepts, Technologies and Applications, Springer, 2007.

[17] “Protégé” [Online]. Available: http://protege.stanford.edu. [Accessed in 2012]

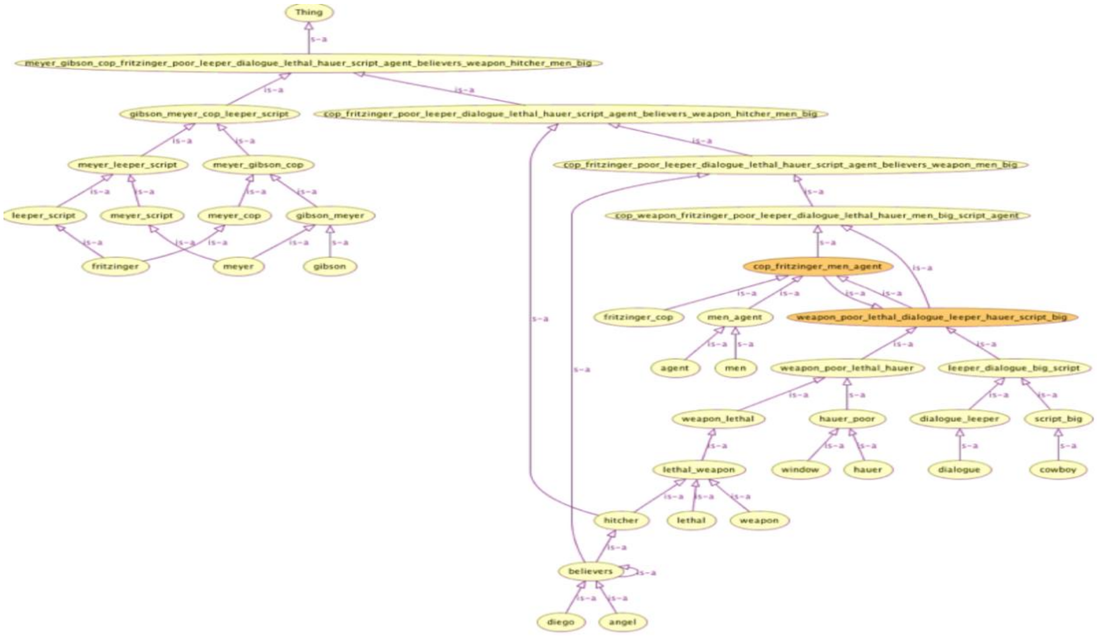

Thing
is-a
meyer_gibson_cop_fritzinger_poor_leeper_dialogue_lethal_hauer_script_agent_believers_weapon_hitcher_men_big
gibson_meyer_cop_leeper_script
cop_fritzinger_poor_leeper_dialogue_lethal_hauer_script_agent_believers_weapon_hitcher_men_big
meyer_leeper_script
meyer_gibson_cop
cop_fritzinger_poor_leeper_dialogue_lethal_hauer_script_agent_believers_weapon_men_big
leeper_script
meyer_script
meyer_cop
gibson_meyer
cop_weapon_fritzinger_poor_leeper_dialogue_lethal_hauer_men_big_script_agent
fritzinger
meyer
gibson
cop_fritzinger_men_agent
fritzinger_cop
men_agent
weapon_poor_lethal_dialogue_leeper_hauer_script_big
agent
men
weapon_poor_lethal_hauer
leeper_dialogue_big_script
weapon_lethal
hauer_poor
dialogue_leeper
script_big
lethal_weapon
window
hauer
dialogue
cowboy
hitcher
lethal
weapon
believers
diego
angel

Figure 13. Generated ontology for the collection of fifteen text-only documents.

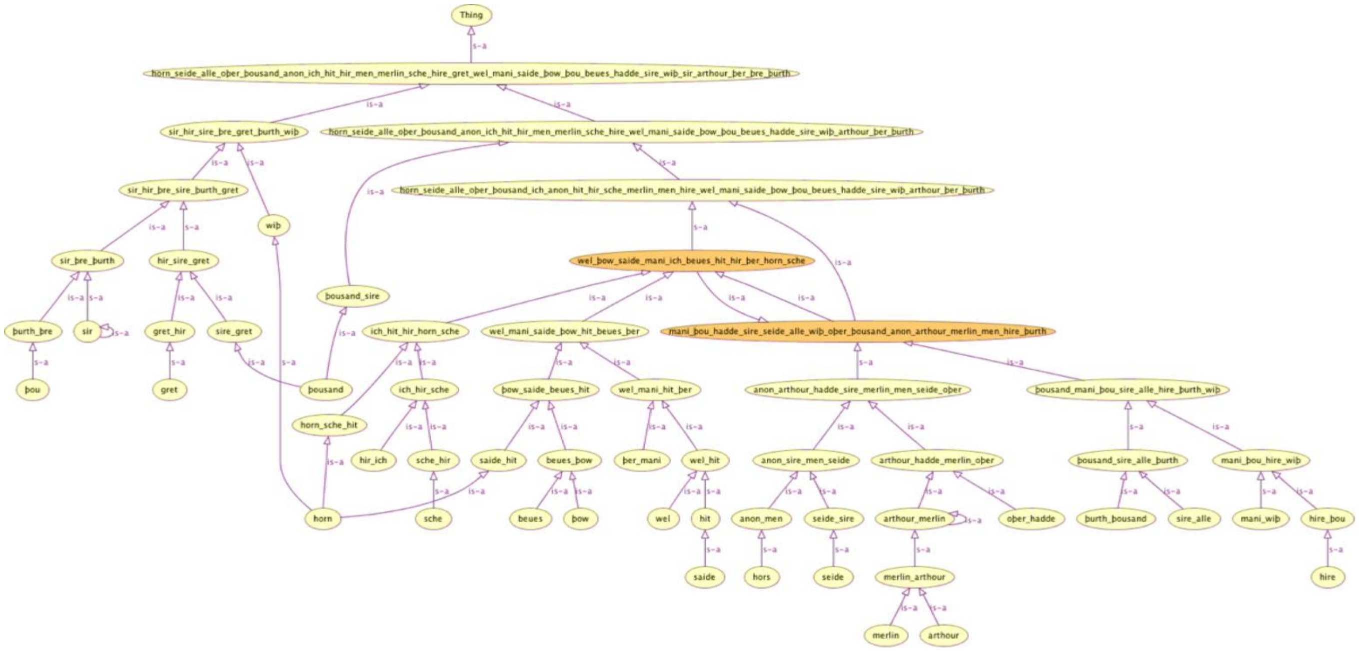

Thing
is-a
horn_seide_alle_oþer_þousand_anon_ich_hit_hir_men_merlin_sche_hire_gret_wel_mani_saide_þow_þou_beues_hadde_sire_wiþ_sir_arthour_þer_þre_þurth
sir_hir_sire_þre_gret_þurth_wiþ
horn_seide_alle_oþer_þousand_anon_ich_hit_hir_men_merlin_sche_hire_wel_mani_saide_þow_þou_beues_hadde_sire_wiþ_arthour_þer_þurth
sir_hir_þre_sire_þurth_gret
wiþ
horn_seide_alle_oþer_þousand_ich_anon_hit_hir_sche_merlin_men_hire_wel_mani_saide_þow_þou_beues_hadde_sire_wiþ_arthour_þer_þurth
sir_þre_þurth
hir_sire_gret
wel_þow_saide_mani_ich_beues_hit_hir_þer_horn_sche
þousand_sire
þurth_þre
sir
gret_hir
sire_gret
ich_hit_hir_horn_sche
wel_mani_saide_þow_hit_beues_þer
mani_þou_hadde_sire_seide_alle_wiþ_oþer_þousand_anon_arthour_merlin_men_hire_þurth
þou
gret
þousand
ich_hir_sche
þow_saide_beues_hit
wel_mani_hit_þer
anon_arthour_hadde_sire_merlin_men_seide_oþer
þousand_mani_þou_sire_alle_hire_þurth_wiþ
horn_sche_hit
hir_ich
sche_hir
saide_hit
beues_þow
þer_mani
wel_hit
anon_sire_men_seide
arthour_hadde_merlin_oþer
þousand_sire_alle_þurth
mani_þou_hire_wiþ
horn
sche
beues
þow
wel
hit
anon_men
seide_sire
arthour_merlin
oþer_hadde
þurth_þousand
sire_alle
mani_wiþ
hire_þou
saide
hors
seide
merlin_arthour
hire
merlin
arthour

Figure14. The ontology representing the whole experiment collection of XML documents.

Facebook as a tool to Enhance Team Based Learning

Sami M. Alhomod
King Saud University, Riyadh-11321,
Saudi Arabia

Mohd Mudasir Shafi
King Saud University, Riyadh-11321,
Saudi Arabia

***Abstract*-A growing number of educators are using social networking sites (SNS) to communicate with their students. Facebook is one such example which is widely used by students and educators. Facebook has been recently used by many educational institutions but most of these have been related to provide the information to the general audience. There has not been much study done to propose Facebook as an educational tool in a classroom scenario. The aim of this paper is to propose the idea of using Facebook in team based learning (TBL) scenario. The paper demonstrates the use of Facebook at each level of TBL The paper shows how Facebook can be used by students and teacher to communicate with each other in a TBL system. The paper also explains teacher – team and teacher – student communication via Facebook.**

Keywords-Social Networking; Facebook; Team Based Learning; Communication.

I. INTRODUCTION

Team based learning (TBL) is based on the use of small groups in order to transform them into high performance teams to accomplish complex tasks. According to Fink [2], "*Team based learning is a particular instructional strategy that is designed to (a) Support the development of high performance learning teams. (b) Provide opportunities' for these teams to engage in significant learning tasks*". There have been a lot of studies that prove that team based learning and teaching have extremely effective to achieve wide range of goals. TBL promotes higher level reasoning, enhances content retention and learning and increases the social support in the classroom. TBL offers an opportunity for an average student to put more effort and enables teams to accomplish tasks which could not have been done by even the excellent students individually [2, 23].

Social networking site like Facebook has gained extreme popularity among the internet user over the past few years. These sites were intended for personal communication among individuals but now increasing number of organizations are using these sites to engage their stakeholders [13]. Facebook is one such site which has seen huge growth since its launch. Facebook offers a means of informal communication among its users [15]. A lot of studies have been conducted recently on the use of Facebook in educational sector. These studies have established that Facebook can act as a tool of communication in the modern educational system. Today more and more educational institutions are experimenting with Facebook as a tool in education.

This paper examines the use of Facebook in a Team based learning system and points out the benefits of using Facebook in a TBL system. The paper is organized as follows; first we will provide a brief background of TBL and Facebook. Second, we will discuss the related work done in this regard. Next, we will describe how Facebook can assist TBL. And at last, we will provide a conclusion to the study.

II. BACKGROUND

A .Team based learning

The term "Team based learning" was first coined by during the 1970's.Team based learning(TBL) in education is a technique in which students work together in teams in order to learn things with better understanding. TBL transforms the traditional lecture based coursework into a more active self-learning and promotes teamwork. It allows students to achieve the levels of higher quality learning which can be hard to achieve when students are working individually [1]. TBL involves making small groups of students and using these groups as instructional strategy. TBL links each learning activity to the next activity in order to achieve deeper understanding among students and develop the teams of higher performance and understanding [2]. As an example, students can be asked to work in teams so that they can cover a more learning material without having to exert excessive pressure individually [3]. According to Michaelsen, Knight and Fink, 2002, there are two specific purposes of TBL:

1. Form Teams of high learning performance.
2. Participate and gain experience in tasks of educational importance.

Another important factor of TBL is group cohesiveness. As the students start working in teams, the group cohesiveness increases which results in higher level of efficiency and understanding among students. Once a student group is formed there are four stages of transforming it into a team. First the students interact with each other. Second, the students review the resources that are available to them. Third students receive a task and work towards its completion. And at last, performance of each individual member of group is evaluated. Once these stages are completed, the group has transformed into the team [1, 4].

One of the important benefits of team based learning is that it helps students with developing skills. Possessing excellent teamwork skills is one of the important factors for employers in the job market [5]. According to a survey conducted by *Wall Street Journal*, a teamwork skill is the second most important skill for the business graduates to possess [6]. TBL allows students to organize the problems and devise a solution for each problem accordingly.TBL also allows students to interact with each other on a daily basis and enables students to complete tasks within teams [4].

B. FACEBOOK

Facebook is one of the most popular Social networking Site today. It was created by Mark Zuckerberg in 2004 while he was a student at Harvard.[9,10] It was initially established to be used by Harvard students only and was extended to be used by other universities and school students in 2005. It was eventually opened to public in 2006. [11,12]. Facebook is currently one of the most sites visited on internet with more than 845 million monthly active users and is available in more than 70 languages.

Facebook is now used by a wide range of people at all levels of education and society as well as by large companies and universities.

Facebook provides its users with numerous ways to communicate with each other. As far as education field is concerned, Facebook can be used to:

a.) Create a group or network
b.) Build educational application for Facebook
c.) Integrate Facebook into current educational tools
d.) Develop a educational application with Facebook developer platform
e.) Share classroom and topic information with other users of the Facebook

According to Withall, (2005) "Facebook.com has become our social Bible for definitive information on our classmates crushes and high school peers we have not spoken to in who-knows-how long." Users use Facebook for daily chat or to discuss events in their life. People also use Facebook to share their thoughts as well as share information and URL's. People also use Facebook to comment on current events and also on some news items.

III. Research Objectives

The trend of using Facebook in a classroom is relatively new and little research has been done with reference to using it in classroom scenario. The aim of this study is to propose the use of Facebook in a team based learning scenario in a classroom. The study proposes following questions:

- How can Facebook be used by faculty in a TBL scenario?
- How can Facebook be used by students in a TBL Scenario?
- How can Facebook assist the Overall TBL Scenario?

IV. Related Work

Facebook has been subject of research among many scholars recently. Facebook has become a big part of student lives and that's the reason Facebook can have a major impact on education. Previous research has established that the university campuses have more than 90% usage rate of Facebook [14,15].

A study conducted by Nemec et.al demonstrated the use of Facebook applications as well as Facebook groups and chats to be used in courses. They concluded that Facebook applications and features can act as a supplement for the classroom program. Another study conducted by Ractham & Firpo, 2011, noted that Facebook allows students to gain informal learning through informal communication, get feedback on their thoughts and communicate without time ans space limit. They also noted the facility provided by Facebook to faculty which allows receiving feedback from students and constantly in touch with students.

Yu-ching Chen found that Facebook provided a common interaction environment which helps in problem solving and reasoning of the people which was reflected by the user satisfaction of Facebook and better academic performance of the students. The study also found that students found Facebook easy, convient, user friendly and simple for academic discussion. Another study by Madgea et.al proposed that Facebook can act as an important tool to aid students in settling in the classroom. They also proposed that Facebook can enhance their team work and organizational skills.

Mazer et.al conducted a study that found that students achieved higher level of motivation after they accessed the teacher's Facebook page them to create a positive atmosphere in the classroom. Laurie & Paula, 2007 proposed the use of Facebook to promote online library services and events.

V. Preparing For The Facebook Based Tbl

Before the Facebook is introduced in the class, some actions must be done by the teacher. The following things must be done before Facebook is integrated in a classroom:

A. Profile Page: A teacher must create a separate profile page that should only to be used for communication with his students.
B. Create class group page: A separate group page for each course should be created. On this page students can find their classmates and can communicate with them as well as with the teacher.
C. Create a team group: Each team must create a team group page for discussion with the other members of the team.

Once the Facebook environment for TBL is created a teacher can continue with the integration of Facebook into the courses.

VI. Facebook Based Team Based Learning

In TBL majority of class time is spent on activities so that students can learn to solve problems which they are likely to face in professional world. According to Michaelsen et al [20], there are three phases of any Team based Learning; preparation phase, application phase, assessment phase. We will discuss how Facebook can assist and enhance each of these three phases [21].

A. Facebook based preparation phase

In this phase students read the topics before they are discussed in the class. The main aim of this activity is to have a prior knowledge about the topic to be discussed in the class. This phase starts with individual preparation of the topic by each student of the group followed by discussing the topic in the group. The students first undergo the test individually followed by the same test in the group based on the topic. Both the tests are graded in class and announced. At the end teacher offers understanding of the concepts that were not understood by the student. This marks the end of the preparation phase.

Facebook can be of extreme importance in this phase. Firstly, students can communicate with teacher as well as among themselves to know about the topics to be discussed in the class before they actually enter the class over the classroom group page. The teacher can share the topic name on their classroom group page and all the students can have a prior knowledge about the topic. The teacher can also choose to give individual task to teams by sharing the topic on the team's group page. Once the topic is provided, the students in the group can discuss topic among each other both prior to coming in the class and during the class. Students in the group can communicate with each other by constantly sharing their thoughts on the question. This will enable students to work individually on the question and discuss it in the group. This will also help teacher to know about the students actively trying to solve the question and also about the students who are not contributing to solve the problem. This may help teachers to identify the weaker student in the group and possibly put more effort towards the weaker students of the group.

Once the tests are conducted individually as well as in teams, the teacher can declare results and provide solution to the question over the classroom group page or by posting the solution to the each team page. This will allow each student to cross examine their result and check the weakness in their answer. Once the results are announced the students can discuss the results among themselves. After the class is over the students can continue discussing the question and can continue discussing the topic even at their home. Using Facebook in this way enables to continue the classroom activity even after the class. The teacher can also participate in the discussion even after the school is over. The teacher can add to the topic at any time. The teacher can choose to inform few things about the topic so as to start the discussion and then gradually add to the topic. This will increase better reasoning among students and can help students to research more about the topic.

Facebook can also allow students an easier way to ask questions. If there is any point in the topic that a student don't understand or need more clarification, they can ask teacher or his team by raising the point on Facebook at any time.

B. Facebook based Application phase

In this phase students apply the knowledge of the course content they learned during the preparation phase to solve the problems, make predictions or create explanations for complex problems. Each group or team in this phase provides their responses to the problem in the class and the teacher evaluates the responses of each group to provide feedback to every group. At the end of this phase, students learn to work in team to provide solutions as well as form a strong bonding with the other students in the group.

Facebook can help in this phase by connecting students with each other as well as with teacher. The students and teacher can communicate with each other over the team specific group page. The constant contact between students and the teacher create cohesion among them which is important for student persistence [14]. Facebook can act as a constant medium for student groups and teacher interaction which is an important factor for the success of students [22].

The students in this phase can solve the problems and share their responses on their team page with other members of the team and with teacher. The teacher can post their responses on the Facebook page of the team. This will create a secret form of communication between teacher and the students. The teacher can choose to guide a particular group or student if they are not doing well as compared to the class. The student can post their arguments and question anytime on the team or classroom page and the teacher can choose to respond to these questions at any time. This will promote out of class learning and students and enable anytime / anywhere learning.

C. Facebook based Assessment phase

This is the Final phase of team based learning. In this phase teams are required to solve the problems based on the understanding of the course material [21]. This phase also allows students to use the previous studied material and incorporate it with the new material [1]. The responses from each team are evaluated by the teacher and the grades for each student and teams are decided.

Timely feedback is one of the fundamental principles of team based learning. It helps students in content retention and learning which in turn helps in student and team development [2]. Facebook can help in this notion of timely feedback. A teacher can provide feedback to student and teams on Facebook as soon as he is done with the assessment. For student it provides an opportunity to readily assess their performance as well as the performance of the team. A team can also share their comments regarding their results and performance.

Besides helping teacher and teams to accomplish their tasks, Facebook can also help in the formation of teams. A teacher can look at the profile page of students and can form teams based on students with similar interests. Facebook can also act as a debate starter. For example, a teacher can post some topic on classroom group page and ask teams about their thoughts about the topic. This can encourage out of class learning. Teams can discuss the topic over Facebook before the topic is actually discussed in the class.

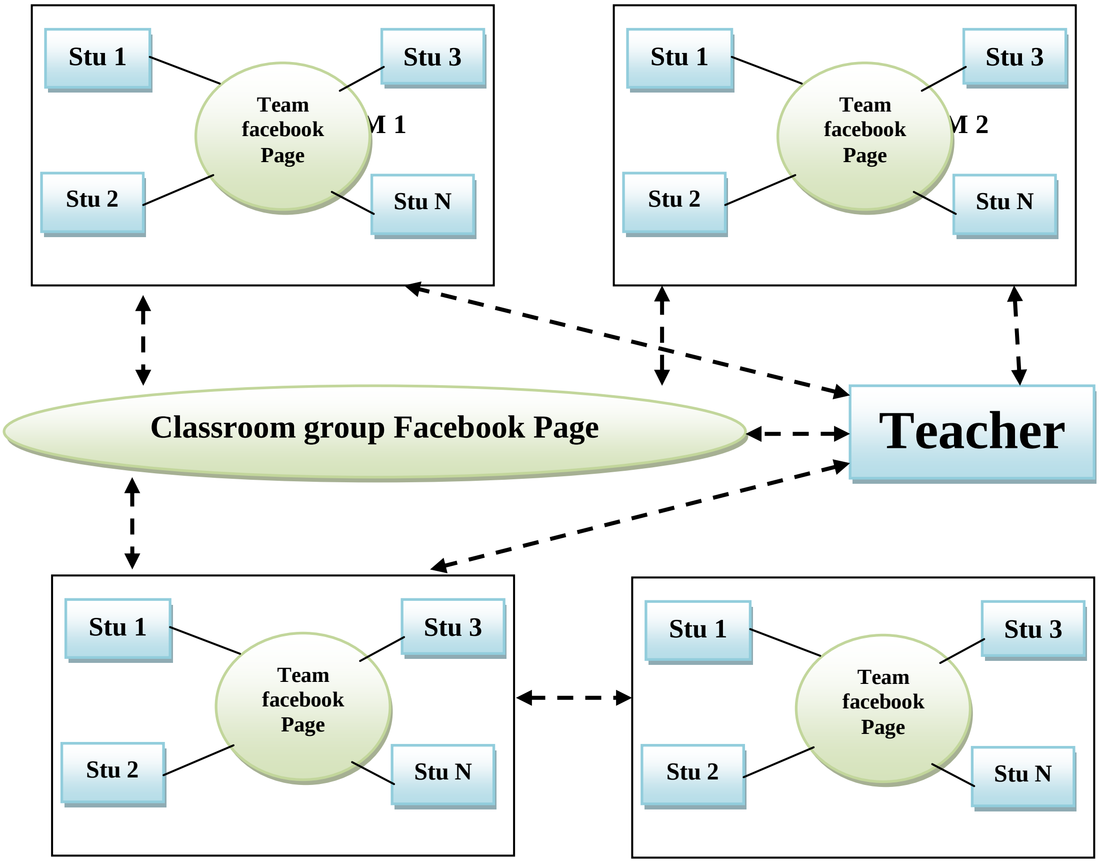


Fig. Facebook based TBL

Fig. above shows a general framework of using Facebook in a classroom structure. A team can communicate with each other using Facebook as well as communicate with teacher and other students in class via Facebook. Teams can set up their intra team Facebook page and discuss the problem assigned to the in team. Teams can remain in constant touch with each other even outside the class and can increase cohesiveness among the team which is important for the successful completion of the tasks. Students can also communicate individually with groups, teacher and other students via a common classroom Facebook page. Teachers can disseminate classroom information on the common Facebook page of the class. As far as teacher- team communication is concerned, this can be accomplished by forming a team specific page. Each team in the class can for a page which is connected only to their teacher. Teams can post their questions on this page and can receive answers from teacher on this page without the interference of out of team members. This can help to increase the teacher- team communication and can allow teacher to access the progress of each individual team. If the teacher has comments for a particular team, teacher can post his comments on the team specific page without other teams and students getting to know about the comments. This can encourage the teams to work hard without going through embarrassment in front of the class.

VII. Conclusion and Future Work

It is established that TBL can enhance education in multiple ways. Recently Facebook has been used by many educational institutions as a tool to enable student achieve the desired outcomes. More and more educational organizations are using Facebook in one way or the other. Keeping this in mind, we tried to demonstrate how Facebook can further enhance the widely accepted mode of learning i.e. TBL. We tried to establish the high level of inter student, intra team interaction, student – teacher and team – teacher interaction via Facebook. This study demonstrated to use Facebook at each phase of TBL as well as tried to demonstrate the use of Facebook in a general classroom scenario. The paper tried to suggest the ways in which Facebook can be used in a TBL. As the use of Facebook in educational institutions grows, we recommend measuring the impact of Facebook in a TBL scenario. We also recommend measuring the impact of Facebook with other learning scenario as well as explore new ways of integrating Facebook into classroom structure.

REFERENCES

[1] Michaelsen, L. K, Knight, A.B., & Fink, L.D. (Ed.). (2002). "Team-Based Learning: A Transformative Use of Small Groups". Westport, CT: Praeger Publishers.

[2] Larry K. Michaelsen , Dean X. Parmelee , Kathryn K. McMahon, Ruth E. Levine, Diane M. Billings "Team-Based Learning for Health Professions Education: A Guide to Using Small Groups for Improving Learning"

[3] Lerner, L. D. (1995). "Making Student Groups Work". Journal of Management Education, 19(1), 123-125.

[4] Allison Brittney Goo, 2011"Team-based Learning and Social Loafing in Higher Education" Online Available: http://trace.tennessee.edu/utk_chanhonoproj/1423/

[5] Chapman, K. J., Meuter, M., Toy, D., & Wright, L. (2006). "Can't We Pick our Own Groups? The Influence of Group Selection Method on Group Dynamics and Outcomes". Journal of Management Education, 30(4), 557-569.

[6] Alsop, R. (2004, September 22). "How to get hired", Wall Street Journal. p.R8

[7] Pavica Sheldon, "Student Favorite: Facebook and Motives for its Use"

[8] Taylor, Chris (June 27, 2011). "Social networking 'utopia' isn't coming". CNN. Retrieved December 14, 2011.

[9] A.Withall, R. (18 November 2005). Facing the facts about Facebook. The Villanovan.

[10] M.D. Roblyera, , , Michelle McDanielb, 1, , Marsena Webbc, 2, , James Hermand, 3, , James Vince Wittye "Findings on Facebook in higher education: A comparison of college faculty and student uses and perceptions of social networking sites" The Internet and Higher Education, 2010

[11] Lili Nemec, Marko Hölbl, Jernej Burkeljca, Tatjana Welzer "Facebook as a teaching tool"

[12] M. A. Urista, et al., "Explaining Why Young Adults Use MySpace andFacebook Through Uses and Gratifications Theory," HumanCommunication, vol. 12, pp. 215 - 229, 2008.

[13] Newsroom http://newsroom.fb.com

[14] Lee, M. J. W., and C. McLoughlin, "Harnessing the affordances of Web 2.0 and social software tools: can we finally make "student-centered" learning a reality?" presented at the World Conference on Educational Multimedia, Hypermedia and Telecommunications, Vienna, Austria, 2008.

[15] Leutner, D., and J. L. Plass, Measuring learning styles with questionnaires versus direct observation of preferential choice behaviour in authentic learning situations: the visualizer/verbalizer behavior observation Scale (VV-Bos). Computers in Human Behaviour, vol. 14, pp. 543-557, 1998

[16] Peter Ractham Daniel Firpo "Using Social Networking Technology to Enhance Learning in HigherEducation: A Case Study using Facebook "Proceedings of the 44th Hawaii International Conference on System Sciences – 2011

[17] Yu-ching Chen "Learning Styles and Adopting Facebook Technology"

[18] Clare Madgea, Julia Meekb, Jane Wellens c and Tristram Hooleyd "Facebook, social integration and informal learning at university: 'It is more for socialising and talking to friends about work than for actually doing work'"Learning, Media and Technology Volume 34, Issue 2, 2009

[19] Joseph P. Mazer*, Richard E. Murphy & Cheri J. Simonds "I'll See You On "Facebook": The Effects of Computer-Mediated Teacher Self-Disclosure on Student Motivation, Affective Learning, and Classroom Climate"Communication Education

[20] Michaelsen, L. K., Knight, A. B., & Fink, L. D. (Eds.). (2004). "Team-based learning" Sterling, VA: Stylus.

[21] Haberyan, April, 2007 "*Team-based learning in an Industrial/Organizational Psychology course"* North American Journal of Psychology Volume: 9 Source Issue: 1

[22]Kuh G.D. (2002) "*The National Survey of Student Engagement: Conceptual Framework andOverview of Psychometric Properties. Center for Postsecondary Research*",Indiana university, Bloomington.

Availableat:http://nsse.iub.edu/pdf/psychometric_framework_2002.pdf

[23] McKeachie, W., (1999). "*Teaching tips: Strategies, Research, and Theory for College and University Teacher (10th ed)".* New York: Houghton Mifflin.

[24] CHARNIGO Laurie , BARNETT-ELLIS Paula "*Checking Out Facebook.com: The impact of a digital trend on academic libraries"* Information technology and libraries,2007

AUTHORS PROFILE

Dr. Sami M. A. Al Homod is a faculty member in the department of Management Information Systems; College of Business Administration in King Saud University. He worked as the Dean of the eLearning and Distance Learning Deanship. He received his PhD in Information Technology from George Mason University; School of Information Technology and Engineering. He received his Master of Science in Information Systems from George Mason University; Information System and Software Engineering Department. He got his B.S. in Computer Information Systems from King Saud University; College of Computer and Information Sciences. He was A Board member of the Saudi Computer Society. He got a Distinguish Degree in Information Systems from George Mason University and Honor Degree from King Saud University. He is a member of many scientific and administrative committees and attended many conferences and scientific seminars.

Mohd Mudasir Shafi was born and raised in Srinagar, Kashmir, India. He has received his Master of Science in Computer Science from Jamia Hamdard (Hamdard University), New Delhi, India in the year 2009. He is currently working as a Researcher in Deanship of distance and Electronic Learning at King Saud University, Kingdom of Saudi Arabia. He has published many research papers in National and International journals. He has also actively attended many international conferences. His areas of interests are in E governance, Mobile governance, Network Privacy and security, Software Engineering, Social Media and E Learning. Apart from that he has previously worked as Quality Analyst and software engineer at various MNC's in India.

Comparaison between MPPT P&O and MPPT Fuzzy Controls in Optimizing the Photovoltaic Generator

Messaouda AZZOUZI
Faculty of Sciences and Technology
Ziane Achour University of Djelfa
Djelfa, Algeria

Abstract—**This paper presents a comparative study between two control methods in order to optimize the efficiency of the solar generator. The simulation had been established by using Matlab/Simulink software to apply the MPPT P&O and MPPT Fuzzy controls on this system which is supplied through a Boost converter.Many results have been illustrated under standard and then variable weather conditions such as the illumination and the temperature. The voltage and the power of the panel and the battery as well as the duty cycle are well presented and analyzed for the two control methods. The obtained results show the effectiveness of MPPT Fuzzy controller in optimizing the PV generator. These results can encourage the use of this control strategy on solar panels in real time to optimize their yield.**

Keywords-solar energy; photovoltaic; PV; MPPT; P&O; Boost converter; fuzzy; optimization.

I. Introduction

The photovoltaic solar energy is among the renewable energies which have the largest development potential. Photovoltaic (PV) generator is based on the smallest unit which is the solar cell. This last is PN junction that generates electricity when it is exposed to light [1].

There are several circuit models for a PV cell but the Single-Diode model is most used because it is the simplified one. Fig. 1 shows a Single-Diode equivalent circuit of solar cell [3] [4].

The output current of the solar cell is given by:

$$I = I\pi\eta - I\delta - IP\pi. \quad (1)$$

By considering the electrical characteristics of the PN junction, this current can be given by:

$$I = I\pi\eta - I0\ (\varepsilon(\theta(\varsigma+IP\sigma\ )/AKT) - 1) - (\varsigma+IP\sigma)/P\pi. \quad (2)$$

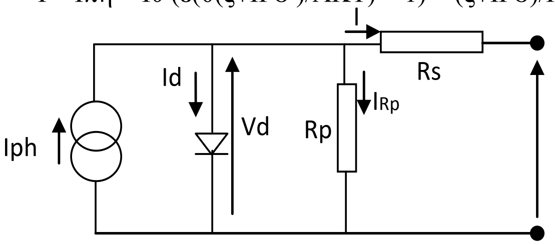


Figure 1. Single-Diode equivalent circuit model of solar cell

When we replace the term $\varsigma_T = KT/\theta$, we find:

$$I = I\pi\eta I0[(\varepsilon\ (\varsigma+IP\sigma\ )/A\varsigma T - 1)]. \quad (3)$$

The output voltage of the cell becomes:

$$\varsigma = -IP\sigma + (AKT/\theta)\lambda\nu((I\pi\eta - I + I0)/I0). \quad (4)$$

The output power of the solar cell is calculated as:

$$\Pi = I.\varsigma. \quad (5)$$

Where:

Pσ: series resistance

Pπ: parallel resistance

Iπη : short circuit current

Iδ: current of the diode

IPπ: current of the parallel resistor Pπ

I: output current and of the solar cell

ς: output voltage of the solar cell

I_0: reverse saturation current of the diode

θ: charge of the electron

A: diode ideality factor

K: Boltzmann constant

T: temperature in ºK

II. Maximum Power Point Tracking

A dynamic tracking method is necessary to extract the maximum power from the PV cells [3]. Many researches has been developed concerning the different algorithms for the maximum power point tracking (MPPT) considering the variations of the system parameters and/or weather changes [2] [6], such as perturb and observe method, open and short circuit method, incremental conductance algorithm, fussy logic and artificial neural network. The block diagram in Fig.2 presents a PV generator with MPPT [5] [11]. The load or the battery can be charged from a PV panel using a MPPT circuit with a specific controller to track the peak power generated by the PV panel.

Other protection devices can be added. The control circuit takes voltage and current feedback from the battery, and generates the duty cycle D, This last defines the output voltage of the Boost converter [10] [13].

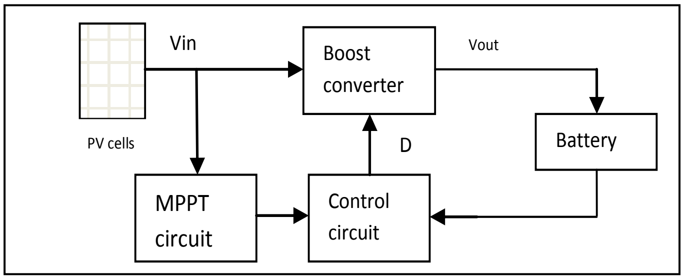


Figure 2. Schematic PV generator with MPPT

A. P&O algorithm

The chart in Fig.3 demonstrates the principle of the Pertub and Observe (P&O) algorithm [5] [7]. This last has been largely used because it is easy to implement, it is based on the perturbation incrementing or decrementing the voltage Vref, or the current Iref with observing the result of this disturbance on the measured power (P = VI) [8] [12].

B. Fuzzy logic

This method uses fuzzy logic to have a faster controller response and to increase system stability once reached the MPP [1]. The tracking of the MPP will be divided into two phases: the first phase is of tough research, with a significant step to improve the response of the MPPT controller, the second one is the fine phase where the step is very small, thus ensuring the system stability and decrease the maximum oscillations around the MPP. This feature of the fuzzy controller demonstrates its effectiveness and makes it among the best MPP tracking devices [9]. The fuzzy controller consists of three blocks: the fuzzification of input variables which is performed in the first block, it allows the passage from the real domain to fuzzy domain. The second block is devoted to inference rules, while the last block is the defuzzification for returning to the real domain. This last operation uses the center of mass to determine the value of the output. Fig.4 shows the basic structure of the used MPPT Fuzzy controller [9].

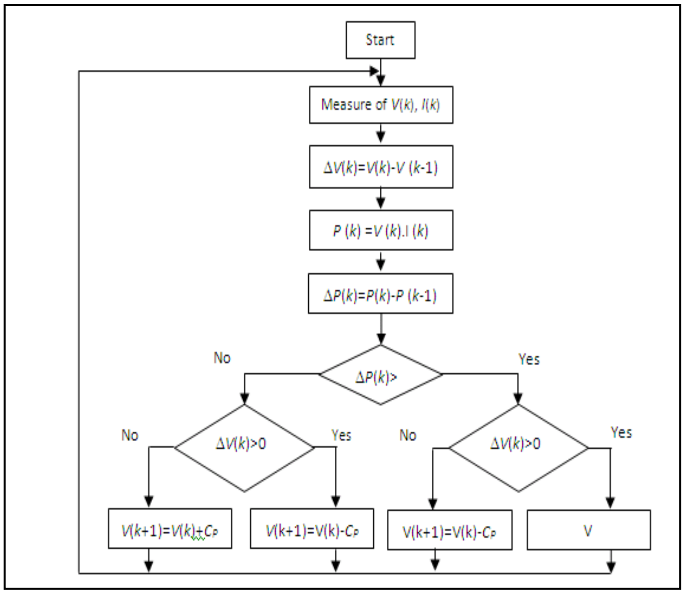


Figure 3. Chart of the algorithm P&O (CP: step width of the disturbance)

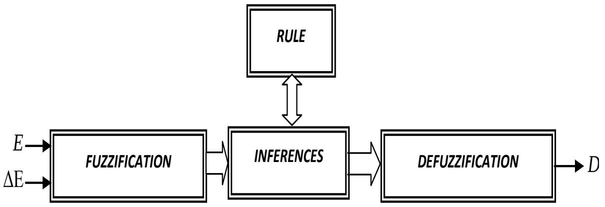


Figure 4. Basic structure of MPPT fuzzy controller

TABLE I. INFERENCES TABLE OF THE FUZZY CONTROLLER

↓ E ΔE →	NG	NP	EZ	PP	PG
NB	EZ	EZ	PG	PG	PG
NS	EZ	EZ	PP	PP	PP
ZE	PP	EZ	EZ	EZ	NP
PS	NP	NS	NS	EZ	EZ
PB	NG	NG	NG	EZ	EZ

The proposed MPPT fuzzy controller has two inputs and one output. The two inputs are the error E and the error variation ⊗E taken at each sampling step κ. These two variables are defined by:

$$E(\kappa) = (\Pi\pi\eta\ (\kappa) - \Pi\pi\eta\ (\kappa - 1))/(\varsigma\pi\eta\ (\kappa) - \varsigma\pi\eta\ (\kappa - 1) \quad (6)$$

$$\otimes E(\kappa) = E(\kappa) - E(\kappa - 1) \quad (7)$$

Where:

$\Pi_{\pi\eta}$: Instantaneous power of the PV generator;

$\varsigma_{\pi\eta}$: Instantaneous voltage of the PV generator.

The value of E(κ) shows the position of the operating point for the load at time k relative to the maximum power point. The value of ⊗E(κ) expresses the direction of movement of this point [1].

The method chosen for inference in our work is that of Mamdani, and for the defuzzification we used the center of gravity method for calculating the output Δ. The duty cycle of DC/DC converter is given by:

$$\Delta = \sum_{j=1}^{n} \square(Dj) - Dj / \sum_{j=1}^{n} \square(Dj) \quad (8)$$

The inference rules can make the right decision for output D from the values of E and ΔE. We chose the rules presented in Table.I.

III. OPERATION IN STANDARD CONDITIONS

The figures below allow us to visualize the variation of the duty cycle and the powers of the module and the battery as the voltages of the module and the battery with P&O and then fuzzy controllers in standard atmospheric conditions (1000W/m^2, 25°C) [14].

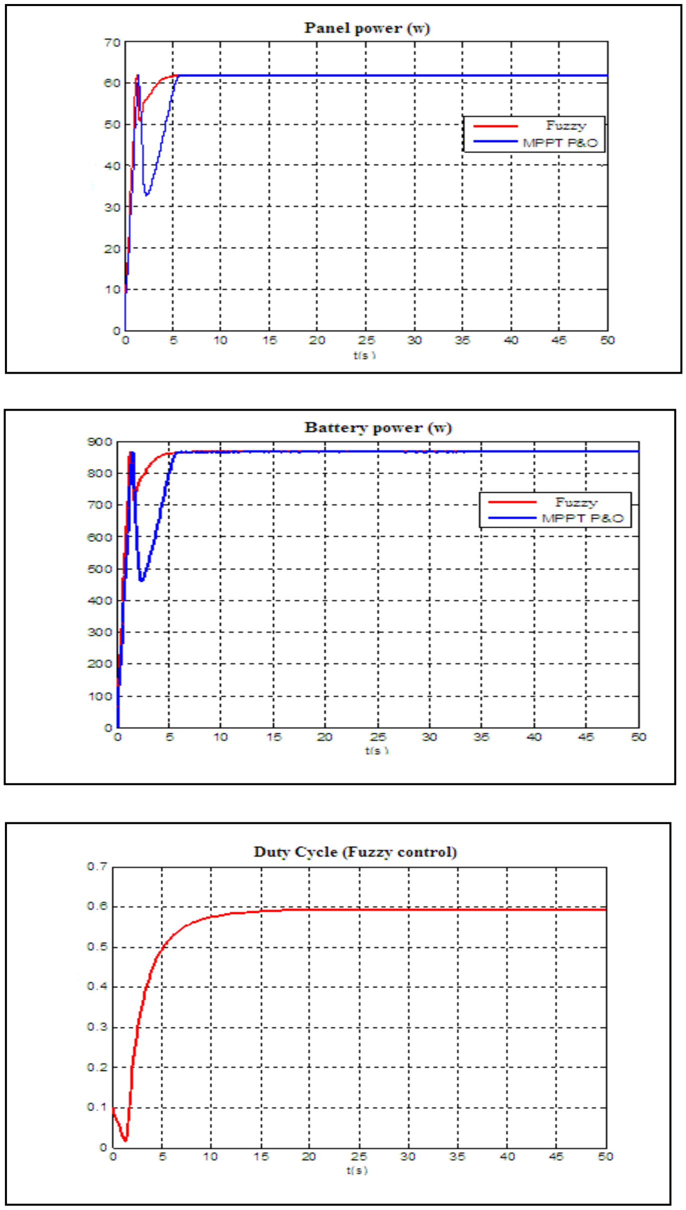
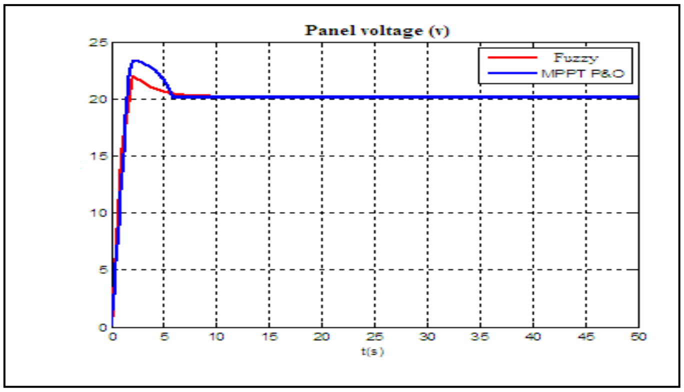


Figure 5. Power variation of the module, and battery and duty cycle D for both controllers and under standard conditions (1000W/m2, 25°C)



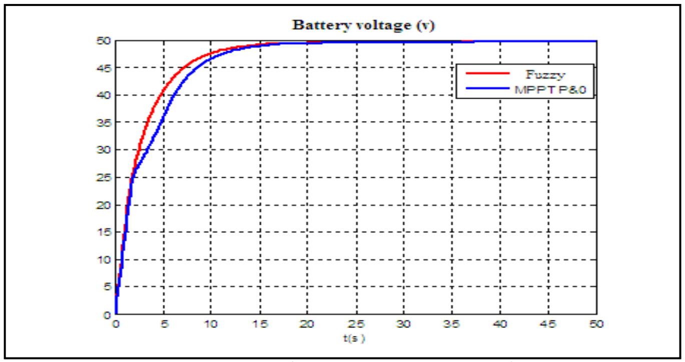


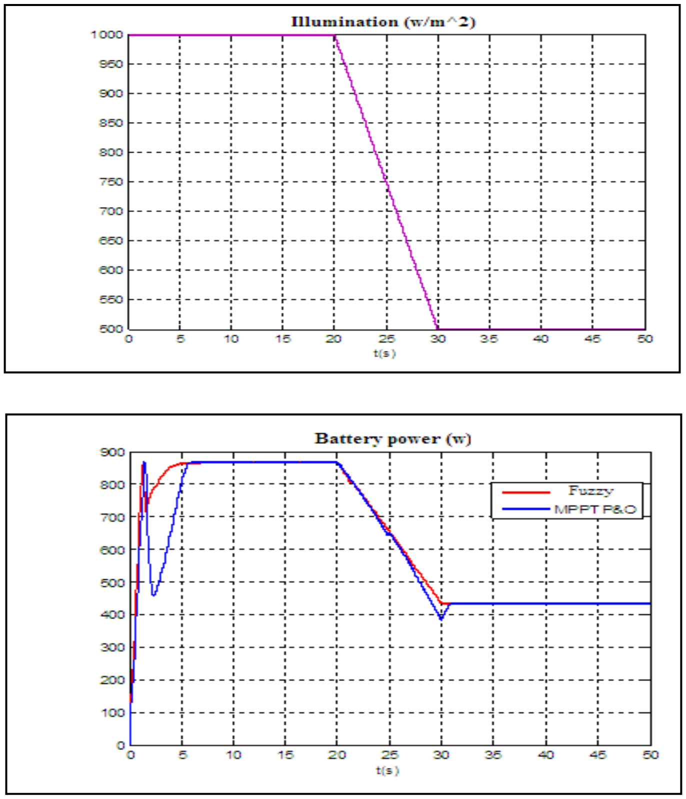
Figure 6. Voltage variation of the module, and battery for both controllers and under standard conditions (1000W/m2, 25°C)

IV. Operation In Variable Conditions

To visualize the behavior of our system in real conditions, we vary the illumination and the temperature, as the increment step. These variations allow us to study the robustness of our system.

A. *Effet of the illumination variation*

In what follows, we will test the response of the two controllers, for a change in illumination from1000 W/m^2 to 500 W/m^2, and this in order to confirm any potential performance presented by this command. The results of simulation illustrated in Fig.7 are considered while the temperature is kept constant throughout the simulation interval at 25°C [14].



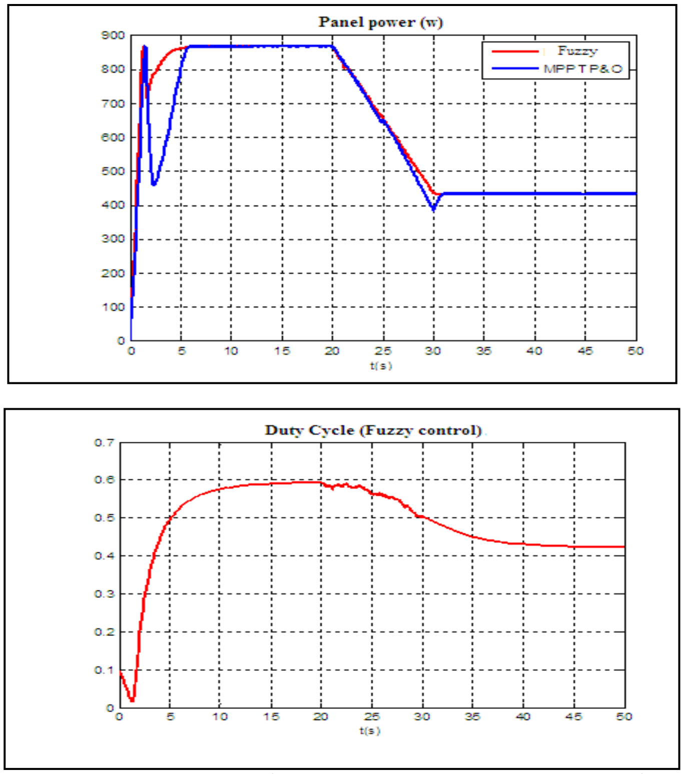


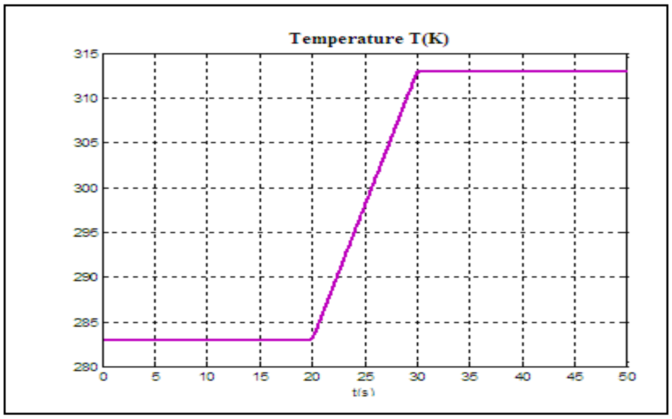
Figure 7. Power variation of the module, and battery and duty cycle D for both controllers for a diminution of the illumination

B. *Effect of temperature variation*

It is very important to test the performance of the command, with respect to possible variations in temperature. It is also considered a state variable whose power PV system depends heavily. The parameter of illumination is kept constant at 1000 W/m^2 for control and during the entire simulation time. The temperature increases from 10°C (283K) to 40°C (313K) (Fig.8) [14].

C. *Effect of simultaneous variation of illumination and temperature*

Fig.9 shows the simultaneous disruption of weather. An increase of the illumination from 500W/m^2 to 1000W/m^2, and temperature from 283K (10°C) to 313K (40°C), with the electrical characteristics of the module and the battery and so the duty cycle [14].



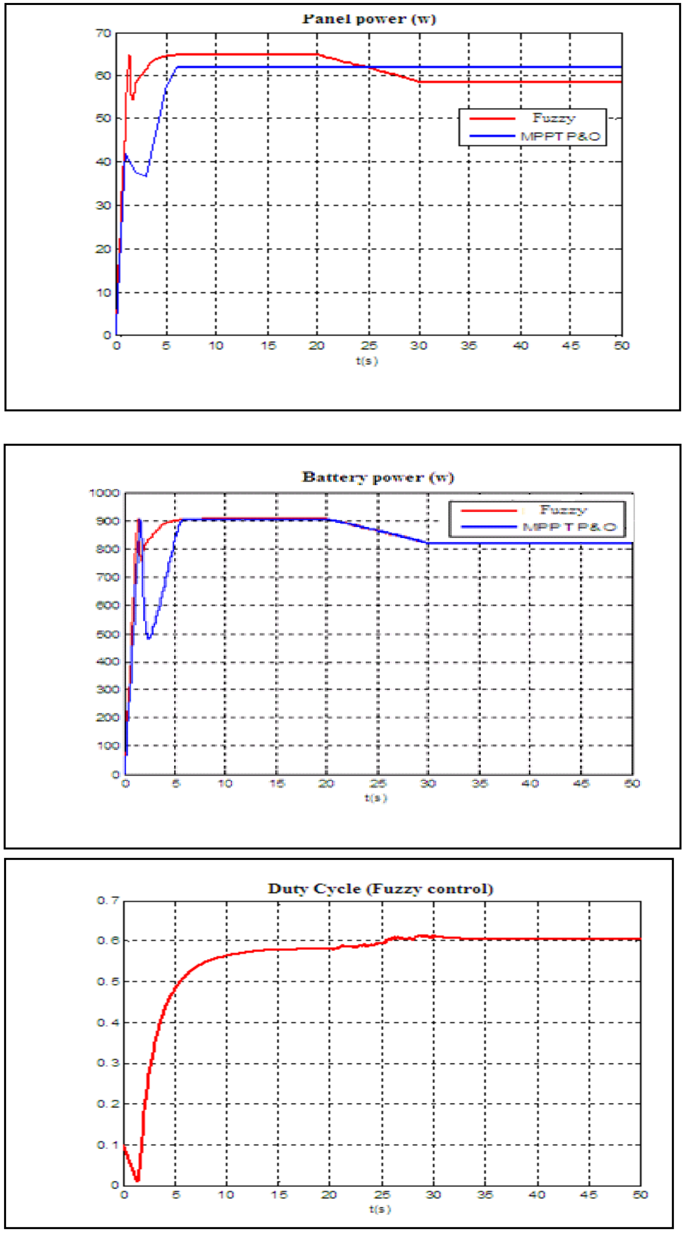
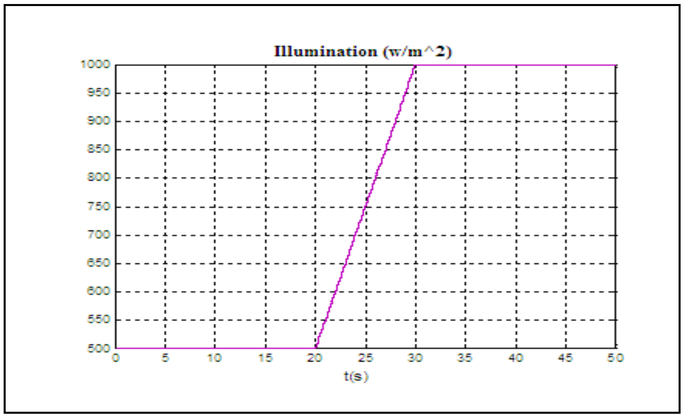


Figure 8. Power variation of the module, and battery and duty cycle D for both controllers for an increase of the temperature



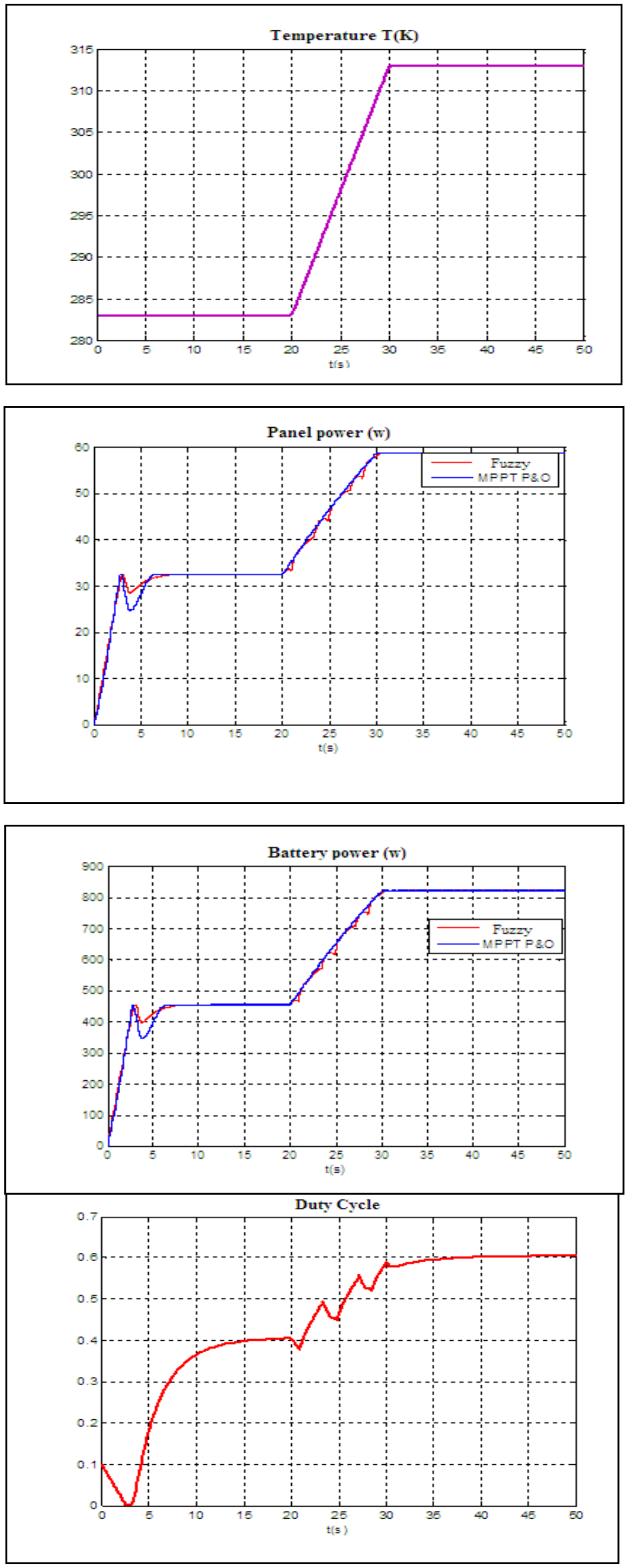


Figure 9. Power variation of the module, and battery and duty cycle D for both controllers for a simultanious increase of illumination and temperature

V. Conclusion

We have seen in this study in detail the simulation of two methods of control: perturb and observe (P&O) and fuzzy controllers, both of them were applied on a chain of energy conversion supplied by Boost converter. We compared the obtained simulation results, by subjecting the controlled system to the same environmental conditions.

We can conclude that MPPT fuzzy controller, is based on the experience of the operator. It has a very good performance. It improves the responses of the photovoltaic system, it not only reduces the time in response to the continued maximum power point but it also eliminates the fluctuations around this point. The fact that shows the effectiveness of fuzzy controller for photovoltaic systems in standard as in variable environmental conditions. The results obtained for this energy conversion system, show that by using the MPPT fuzzy controller, there is a compromise between rapidity in transient regime and stability in steady state.

These used controllers results can be compared to other methods of control as using neural networks in optimizing the photovoltaic generator power, the idea of our future work as extension of our research to improve more the PV systems yield.

References

[1] A. Chouder, F. Guijoan and S. Silvestre, Simulation of fuzzy-based MPP tracker and performance comparison with perturb & observe method, Revue des Energies Renouvelables, Vol 11, No 4, pp.577-586, 2008.

[2] C. Cabal and C. Alonso, Adaptive Digital MPPT Control for Photovoltaic Applications, IEEE Trans Power Electronics, pp. 2414-2419, 2007.

[3] C. L. B. Wu and R. Cheung, Advanced algorithm for MPPT control of photovoltaic systems, Canadian Solar Buildings Conference, Montreal, 2004.

[4] D. Petreuş, C. Fărcaş and I. Ciocan, Modelling and simulation of photovoltaic cells, ACTA TECHNICA NAPOCENSIS- Electronics and Telecommunications, Vol 49, No 1, pp.42-47, 2008.

[5] J. A. Jiang, T. L. Huang, Y. T. Hsiao and C. H. Chen, Tamkang Journal of Science and Engineering, Maximum Power Tracking for Photovoltaic Power Systems, Vol. 8, No 2, pp. 147-153, 2005.

[6] K. Ameur, Étude d’un Système Photovoltaïque Muni d’un Régulateur MPPT : Application à la Conduite d’une Machine Synchrone à Aimants Permanents", Thèse de Magister, ENP, Alger, 2009.

[7] M. A. Elgendy, B. Zahawi and D. J. Atkinson, Assessment of Perturb and Observe MPPT algorithm implementation techniques for PV pumping applications, IEEE transactions on sustainable energy, pp.21-33,Vol 3, No 1, 2012.

[8] M. Angel Cid Pastor, Conception et Realisation de Modules Photovoltaiques Electroniques, These de Doctorat de l’Institut National des Sciences Appliquées de Toulouse, Spécialité: Conception des Circuits Microélectroniques et Microsystèmes, 2006.

[9] M. Hatti, Controleur flou pour la poursuite du point de puissance maximum d’un système photovoltaique, Huitieme Conference des Jeunes Chercheurs en Genie Electrique (JCGE'08), Lyon, 2008.

[10] N. Femia, G. Petrone, Giovanni Spagnuolo, and Massimo Vitelli, Optimization of Perturb and Observe Maximum Power Point Tracking Method, Transactions on power electronics, pp.963-973, Vol 20, No 4, 2005.

[11] P. C. M. Bernardo1, Z. M. A. Peixoto1 and L.V. B. Machado Neto, A High Efficient Micro-controlled Buck Converter with Maximum Power Point Tracking for Photovoltaic Systems, International Conference on Renewable Energies and Power Quality (ICREPQ’09), Valencia, 2009.

[12] R. Faranda and S. Leva, Energy comparison of MPPT techniques for PV Systems, WSEAS transactions on power systems, ISSN: 1790-5060, pp.446-455, Issue 6, Vol 3, 2008.

[13] R. Leyva, C. Alonso, I. Queinnec, A. Cid-pastor, D. Lagrange, l. Martínez-salamero, (2006), MPPT of photovoltaic systems using Extremum–Seeking control, pp.249-258, Vol 42, No 1, IEEE transactions on aerospace and electronic systems.

[14] T. Mrabti, M. El Ouariachi, B. Tidhaf et K. Kassmi, (2009), Caractérisation et modélisation fine du fonctionnement électrique des panneaux photovoltaïques, Revue des Energies Renouvelables Vol. 12 No 3, pp.489-500.

Learning from Expressive Modeling Task

a Mathematical Model by Electronic Spreadsheet for the Car's Trip Computations

Tolga KABACA
Department of Teaching Mathematics
Pamukkale University
Denizli, Turkey

***Abstract*—This study aimed to present an authentic way of showing how computer assisted mathematical modeling of a real world situation helped to understand mystery of that situation. To achieve this aim, a group of pre-service mathematics teachers has been asked to think on how the trip computer of cars calculates the values like instant fuel consumption, average fuel consumption and the distance to be taken with remaining fuel. The theoretical discussion on mathematical structure has been directed as semi-structured interview. Then, theoretical outcomes have been used to create the model on the electronic spreadsheet MS Excel. At the end of the study, it has been observed that students have easily understood the behavior of trip computer by the help of mathematical background of the spreadsheet model and they have also been awaked of the role of mathematics in a real sense.**

Keywords-Computer Assisted Modeling; Electronic Spreadsheet; Mathematical Model.

I. Introduction

After a comprehensive literature synthesis about modeling by using technology, Doer and Pratt propose two kinds of modeling according to the learners' activity; "exploratory modeling" and "expressive modeling" [1]. In exploratory modeling, a learner uses a ready model, which is constructed by an expert. In expressive modeling, he or she shows his/her own performance to construct the model. During the process of constructing the model, learner can find the opportunity to reveal the way of understanding the relationship between the real world and the model world [1]. If we can give an expressive modeling task related with a realistic problem from our real life, this can provide a chance of understanding the real world by mathematics. It will be better to suggest using a well-known technology to our students while studying on their modeling task. This will prevent some unexpected problems about technological tool rather than understanding the problem situation and mathematical activity.

Electronic spreadsheets like MS Excel are good tools while understanding some hidden relations between variables and it is also an easy technology to use for most of the students. Electronic spreadsheets were declared as a practical tool that helps students to focus on mathematical structure of the concepts deeper instead of struggling on complex, difficult and time-consuming operations [2, 3]. Some researchers used spreadsheets to teach some concepts and make them understandable by modeling activities [4, 5, 6].

Grossmann used the spreadsheet modeling method to make the queue behavior understandable in a business school end user modeling course (1999). At the end of the study, Grossmann advocated that spreadsheet modeling simulations are surprisingly easy to program and this method develops the intuition of students. Besides, students find the opportunity of developing their modeling skills. Dede and Argun (2003) emphasized that electronic spreadsheets provide opportunities to make connection between numerical, algebraic and graphical representations of the concepts. Ozmen (2004) used the spreadsheets on investigating the solutions of partial differential equations. Kabaca and Mirasyedioglu (2009) proposed an approach to teach the concept of differential by using MS Excel and they concluded that this numerical approach created more meaningful sense in students' minds.

In this context, this research primarily focused on using electronic spreadsheet for a real life modeling problem and examined the student's thinking and learning process from this modeling task. The modeling task stated as "Can you find an explanation for how the trip computer (TC) of cars works? How does a TC calculate the instant speed, instant fuel consumption, average fuel consumption, average speed and the distance to be taken with remaining fuel?", the secondary purpose of this paper is making students to understand the mathematics' role in the world by using a context which is a mysterious thing for most of the people.

II. Methodology

The task was given to a group of pre-service teachers who are taking an elective project course in a faculty of education in Turkey. The group was containing 4 students and they were asked to finish the task in three weeks. During the working process, the group and the instructor met several times and discussed the progression of the work. Every class administered as a semi structured interview and reported with nicknames of students. These classes were reported as 5 different titles, which reflect corner stones of the modeling task.

1. Initial discussion and determining what we need before starting to work with Excel. In this discussion it is concluded that we need to reach volume of the tank by using its fuel level. Beside this, we also need to discuss some theoretical issues.

2. Designing a sample fuel tank to make the volume computable in terms of the fuel level.

3. Discussion on theoretical structure of the model.

4. Formulizing the electronic spreadsheet using the theoretical structure.

5. Discussion on the reflections of the model to the understanding of the data of the cars and some mathematical concepts.

A. Description of the Modelling Task

A real situation was chosen from the world of cars. The trip computers (TC) which are among the indispensable technologies of our cars in the recent years present us the data as instant or average fuel consumption, distance to be taken by remaining fuel, average speed and travel time by the mediation of a little screen. If you do not have this system in your car, you can calculate the fuel consumption that matches the unit distance you took by using a more conventional method as follows; Fill the fuel tank up to the level it floods. After taking a certain distance, fill your fuel tank again up to same level. After the second filling operation, if you divide the quantity of the fuel that the tank holds by the distance you get between two filling operation, you can calculate how much fuel does your car consume while getting a kilometer distance. Since this value is generally very little, in order to make it more clear by multiplying it by a hundred you can imply in a more clear way how much liter fuel it consumes during a hundred kilometer. TC also presents consuming values in the category of consumption at 100 kilometer.

In this case a question like this may occur in our minds: “if we have the capacity of calculating this data, why the use of TC is needed?” We answer this question in two ways: Firstly, with the method we mentioned above, we can only calculate the fuel consumption between two certain points. If we wonder how much our car consumes at a certain time while we are driving, we need both more information than we mentioned and a more complex calculation. Secondly, it may be a cautionary factor to drive more economically that whenever we look, to be able to check the instant consumption.

III. Findings

Discussion sessions started by deciding what we have in our hands and which data must be calculated in our model.

A. Initial Discussion

Instructor: *As you know, we can easily know how much fuel exist in our car’s tank and how far we go from a specific point, where we restarted our car’s trip measurer. Besides these, we can easily measure the time elapsed. So, we have the following variables; the time, the amount of the fuel and the distance traveled. Can you list the variables that we need to calculate for TC?*

Student-A: *Sorry! How can we know the amount of fuel in our car’s tank?*

Student-B: *All cars can display the fuel level with a fuel gauge!*

Student-A: *Yes I know! But, this is only level. It does not guarantee the exact amount of the fuel in the tank.*

Instructor: *You are right! We also need to calculate the volume of the fuel by using its level in the tank. For a while, assume that you know the amount of the fuel at a specific time and let’s think on how we can evaluate instant and average fuel consumption.*

Student-C: *It is related to lots of variable. I think we can not control everyone at the same time.*

Student-A: *Drive style, weather conditions, quality of the car. This kind of variables effect the fuel consumption. I still think that we can not calculate the consumption.*

Instructor: *Yes! You are right! There are lots of things that possibly effect the consumption. But, all of these have the role on indicating the volume of the fuel tank. We just want to know the result. So you can find a solution for evaluating consumption values by using changing the volume of the fuel tank.*

Student-C: *I think the key word is “change”. If we can obtain the volume of the fuel tank at every specific time, we can control change on volume of the fuel according to the time.*

So far, students reached a valuable result on the way of solving problem. The world “change” hosting the basic mathematical concept that will be useful for the model. On the other hand, we have a new problem of finding a way to evaluate volume of the fuel in the tank, in term of the fuel level.

At this step of the task, the instructor decided to give a sub-problem of creating a virtual fuel tank and calculate its volume in terms of its level.

B. Designing a sample fuel tank

Instructor: *a basic car can indicate the level of the fuel by the help of a gauge. Of course, our modern cars may find a technological way to obtain the volume of the fuel. Now, let’s find the volume mathematically in term of the level. I will give you a model fuel tank and ask you to evaluate its volume in term of its height.*

Student-A (who is more interested in cars): *the change speed of the level is getting faster and faster as coming close to the end of the tank. So, level is not a good indicator to trust.*

Instructor: *Yes! Your friend is completely right! The source of this behavior is the shape of the tank. This is why we need to find the volume instead of the level. The pointer that shows the fuel level declines quickly especially in the last quarter when the fuel is less than half of the tank while it declines slowly at first quarter or when the tank is half. So the shape of our model tank must model this behavior also. It can be considered that a tank as the one in figure-1 will be a good structure by carrying the properties we look for;*

We have a rectangular prism. And we are extracting two quarter sphere like in the shape. So, upper level of our tank will have more fuel according to its lower level. I hope this shape can model a classical fuel tank’s behavior.

Student-B: *I think, now our work is to obtain a relationship between height and volume.*

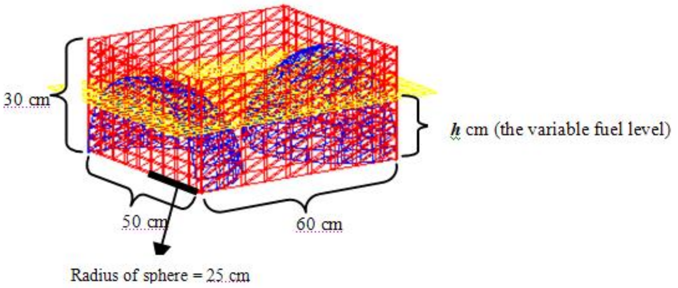


Figure 1. The sample fuel tank

Students worked together and reached following solution under the enough guidance given by the instructor.

Complete volume will be;

$$V_{sum} = 50.60.30\,(\text{rectangular prism}) - \frac{2.\pi.25^3}{3}\,(\text{two quarter-spheres})$$

$$\cong 57275,077\ cm^3 \cong 57,3\,\text{liter}$$

The tank has the volume of an average car. Actually, we need to evaluate the volume as a function of h which means the level of the fuel. According to the figure-1 above, we have two volumes that have different characteristics. At the first volume, the fuel level is changing from 25 cm to 30 cm and at the second volume; the fuel level is lower than 25 cm. let's call the first volume as $V_1(h)$ and it should be defined like in the complete volume evaluation as below;

$$V_1(h\ ) = \frac{50.60.h}{1000} - \frac{2.\pi.25^3}{3000} = 3.h - \frac{125\pi}{12} \qquad 25 \le h \le 30$$

When the fuel level decreases fewer than 25 cm, we should apply double integration to evaluate the inner volume of quarter spheres. Let's just consider on one of the quarter spheres on figure-1. We have to evaluate the volume bounded by the planes y=0, z=h and the surface $x^2 + y^2 + z^2 = 625$ (figure-2).

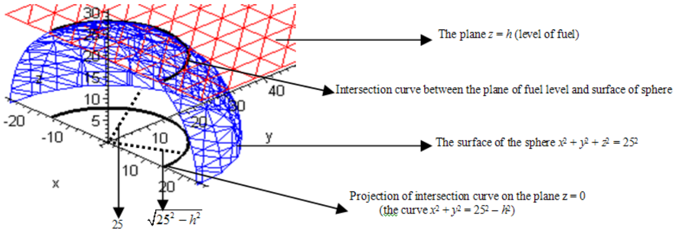

30
25
20
15
10
5
z
-20
-10
10
x
10
20
40
30
20
y
25
$\sqrt{25^2 - h^2}$
The plane $z = h$ (level of fuel)
Intersection curve between the plane of fuel level and surface of sphere
The surface of the sphere $x^2 + y^2 + z^2 = 25^2$
Projection of intersection curve on the plane z = 0
(the curve $x^2 + y^2 = 25^2 - h^2$)

Figure 2. Calculating the quarter sphere part of the tank

According to the figure-2, the desired volume can be determined as following by using cylindrical coordinates.

$$V_{\text{inner sphere}}(h) = \int_{\theta=0}^{\pi}\int_{r=0}^{\sqrt{625-h^2}} hrdrd\theta + \int_{\theta=0}^{\pi}\int_{r=\sqrt{625-h^2}}^{25}\sqrt{625-r^2}\,rdrd\theta$$

$$= \frac{h(625-h^2)\pi}{2} + \frac{h^3\pi}{3}$$

We can reach the second volume $V_2(h)$ by subtracting twice of the volume obtained above. Do not forget that we must multiply by 1/1000 to state the volume as liter.

$$V_2(h) = \frac{50.60.h}{1000} - \frac{2}{1000}\left(\frac{h(625-h^2)\pi}{2} + \frac{h^3\pi}{3}\right)$$

$$= 3.h - \frac{h(625-h^2)\pi}{1000} - \frac{h^3\pi}{1500}$$

At last, we can determine the volume function of fuel level as following piecewise function.

$$V(h) = \begin{cases} V_1(h) & ,h \ge 25 \\ V_2(h) & ,h < 25 \end{cases} =$$

$$\begin{cases} 3h - \dfrac{125\pi}{12} & ,h \ge 25 \\ 3h - \dfrac{h(625-h^2)\pi}{1000} - \dfrac{h^3\pi}{1500} & ,h < 25 \end{cases}$$

Instructor: *Well done! It looks as a good work! You can try to plot the graph of function you obtained and check that our fuel tank can model the behavior of a real consumption.*

Student-A: *We obtained the graph on figure-3. When the volume decreases by equal intervals, level decreases faster and faster.*

Student-C: *Yes! This exactly like a real car's fuel gauge! Our fuel tank is really a good model.*

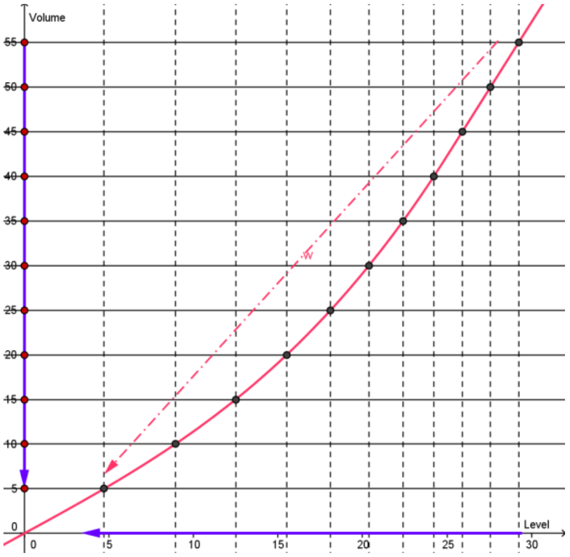


Figure 3. The graph of volume function of fuel level

By using advanced mathematics, students were able to obtain the volume of the fuel in terms of the level. The function that they obtained also has the capacity of modeling the behavior of an ordinary car's level indicator. Now two variables exist. These are the volume of the fuel at a certain time and the distance took by the car from initial time to a certain time.

C. Discussion on theoretical structure

So far, students had a sense about the variables which the car can collect independently. Now, students need to be aware of the dependent variables that TC should compute.

Student-A: *Let's start by studying on calculating instant fuel consumption.*

Instructor: *Make a table including the data, which are collected by car as defined previously.*

Student-C: *I think we need to decide a start point for recording data.*

Student-B: *Yes, this point represents the time that we reset the TC. That is, after a starting time we have the distance took by car and volume of the fuel in the tank.*

Instructor: *Assume that, your car is recording these data by a specific time interval. Let the time be "t_0, t_1, t_2, t_3, ...", distance be "x_0, x_1, x_2 . . ." and the volume of the fuel be "l_0, l_1, l_2 . . ."*

Student-B: *Starting distance x_0 every time must be 0, isn't it?*

Student-A: *Of course, we have a table as below;*

TABLE I. DATA RECORDED BY THE CAR

Time	$t_0, t_1, t_2, t_3, t_4, t_5, \ldots t_{n-1}, t_n, \ldots$
Fuel Volume (liter)	$l_0, l_1, l_2, l_3, l_4, l_5, \ldots l_{n-1}, l_n, \ldots$
Distance (meter)	$x_0, x_1, x_2, x_3, x_4, x_5, \ldots x_{n-1}, x_n, \ldots$

Student-C: *I think the problem is the difference between t_{n-1} and t_n. How much difference is enough for a better evaluation?*

Instructor: *Yes! This is one of the important points for our model. Initially, assume that this interval is 3 second. Your car's computer is recording the data for every 1 second. At the beginning, do not pay attention this issue. Try to think and develop a theoretical structure.*

Student-A: *We need to find a way of evaluation method for instant fuel consumption.*

Student-B: *I think this will be similar with evaluating instant speed.*

The word "instant" evoked the students for instant speed.

Instructor: *The limit of average speed in a time period equals to the instant speed as the time period decreases. We learned this in the Calculus courses. Let's try to apply this concept for instant fuel consumption.*

Student-C: *OK! I remember it. But we do not have any function. How can we evaluate the limit?*

Students remembered the formal way of finding instant speed by using average speed. This maybe said that a concept definition image. Instructor helped students to reconstruct their concept image.

Instructor: *Look at the figure below! Every point x_i represents volume of the fuel at the time t_i and every circle represents the point (x_i, t_i). We know the specific value for each point represented by the circles. It is assumed that these values are measured by the car and it is impossible that this measurement is really continuous. We can only assume that there is a curve connecting these points (represented by dashed curve in the figure). If we knew the algebraic relation of this curve, it would be easy to calculate the limit at a specific t_i of this relation.*

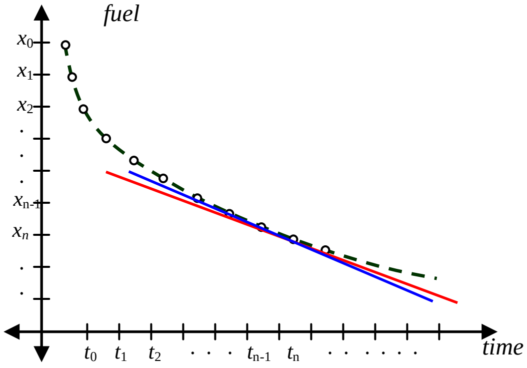


Figure 4. The graph of change of fuel related to time

Student-C: *I also mean that how we can evaluate the limit while we do not know the function.*

Instructor: *Yes! We have discrete points instead of a continuous curve represented by an algebraic relation. So, we have to focus on the background of the concept. You can easily notice that two secants' slopes are approximately same. One of these secants has consecutive points while the other is not. I mean one of the secants is approximately tangent. Of course! This approximation is up to the length of the interval [tn-1, tn].*

Let me explain more mathematically;

Let $x_{n-1} - x_n = \Delta x_n$ and $t_n - t_{n-1} = \Delta t_n$. As we said before, if we knew the algebraic relation, we could find the slope of the tangent, which means instant change rate of fuel, by following operation;

$$\lim_{\Box t_n \to 0} \frac{\Box x_n}{\Box t_n} = \left.\frac{dx}{dt}\right|_{t=t_n}$$

In other words, the derivative of the fuel function of time can help us to find the instant fuel consumption.

Student-A: I got it! But we do not have still the algebraic relation and it is seen impossible.

Student-B: Maybe, we need to use the logic of approximation. But, I do not know how!

Instructor: Well done! Since we do not have the algebraic relation, which provide continuous values for every time, we cannot perform the formal limit operation, which will provide a perfect result.

So, we have to use Δx/Δt, instead of its limit as Δt goes to 0. Of course, we have to make Δt as small as we can measure.

Surely, that is not the only case we must discuss about. Even if we use the term "instant fuel consumption" under this title, our car does not show the quantity of the fuel consumption in unit time but it shows the quantity of the fuel at

100 kilometers that can be consumed during the time we are in. What does this mean? Firstly, the fuel consumption at 1 kilometer according to our driving feature during the time we are in is calculated than it is presented after multiplying by 100 as it is a too low value to reflect on the screen. If you be careful, the data about instant fuel consumption is written on the screen called TC with “x lt/100km” unit. As distance taken and the quantity of the fuel in the tank are unavoidably related to time variable, probably that is why the car firms use “instant fuel consumption” phrase.

This data can be calculated with the help of the values at chart 1. The only thing we should do is that to get benefit from the relationship between “fuel quantity- distance taken” rather than using the “fuel quantity- time” relationship in figure 4. Figure 5 shows this relationship.

After discussing on the above issue with the students, they asked to use relationship between “fuel quantity- distance” as in figure-5 rather than using the “fuel quantity- time” relationship in figure-4.

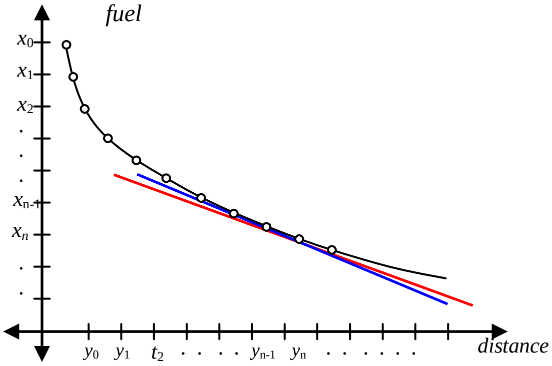


Figure 5. The graph of change of fuel related to distance

Student-C: *Can we say “we will use an approximate derivative instead of the perfect and formal derivative concept”*

Instructor: *Sure! But this is not the only case. You also should state the variable which is independent for the derivative operation.*

Student-C: *The independent variable must be the distance.*

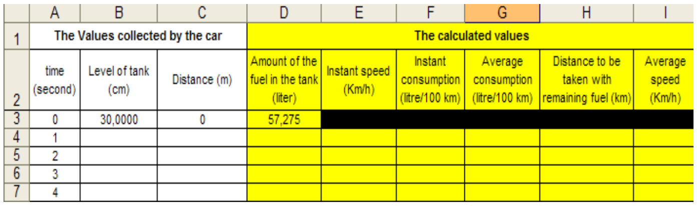
Displaying two lines, which one is exact tangent representing the derivative and the other one is just a secant passing through two close points, helped students to state the phrase of approximate derivative.

D. Programming the electronic spreadsheet

We completed the preparation of the work which was for getting the data that TC present. At the end, we saw that we must apply the same operation regularly on the discrete values for each second.

In order to operate the data at table-1 regularly, it is advised to use an electronic table processor and a ready template is given to the students by asking them to formulate it (Table-2).

TABLE II. ELECTRONIC SPREADSHEET TABLE PREPARATION



	A	B	C	D	E	F	G	H	I
1	The Values collected by the car			The calculated values					
2	time (second)	Level of tank (cm)	Distance (m)	Amount of the fuel in the tank (liter)	Instant speed (Km/h)	Instant consumption (litre/100 km)	Average consumption (litre/100 km)	Distance to be taken with remaining fuel (km)	Average speed (Km/h)
3	0	30,0000	0	57,275					
4	1								
5	2								
6	3								
7	4								

Instructor: *I have prepared a template for the electronic table that you must formulate. I assume that the fuel level is 30 cm at the beginning and the cell D4 is formulated as displaying the volume. Please remember the volume function in terms of level.*

$$V(h)=\begin{cases}3.h-\dfrac{125\pi}{12} & ,h\geq 25\\ 3.h-\dfrac{h(625-h^2)\pi}{1000}-\dfrac{h^3\pi}{1500} & ,h<25\end{cases}$$

The h independent variable of this function is cell B4 according to the electronic table. Accordingly, the formula that should be written in cell D4 must go as follows:

=IF(B4>25;3*B4-125*PI()/12;3*B4-B4*(625-B4^2)*PI()/1000-B4^3*PI()/1500)

Please think on how the cells D4, E4, F4, G4, H4 and I4 at table-2 should be formulated in order to reach the data wanted. After formulating the line 4, it is enough to copy by dragging this line into successive lines.

Student-A: *You wrote on the time column as 0, 1, 2 . . . Do you mean that we will use the differential as 1 second?*

Student-C: *Yeah… I see! Because the difference is just 1 second.*

After observing that how students got aware of the role of the differential and derivative concepts to calculate the TC data, it is just reported the results that they reached after little help, especially on syntax rules of MS Excel.

Instant speed (the cell E4)

Let y demonstrate the distance taken and let t demonstrate time. The instant speed can be written as below where the $t_n - t_{n-1}$ is the most possible lowest value;

$$\text{Instant sepeed}=\frac{y_n-y_{n-1}}{t_n-t_{n-1}}$$

According to the electronic table at table-2, t_1 and t_0 are A4 and A5 cells respectively an y_1 and y_0 are C4 and C3 cells respectively, the formula that should be written in E4 cell must be like below in order to find the value of instant speed at first second in terms of km/h.

=(C4-C3)/(A4-A3)*36/10

Instant fuel consumption (the cell F4)

By regarding the figure 5, the instant fuel consumption, which means the fuel consumption at unit distance instead of unit time actually, can be written as below.

$$\text{Instant Consumption} = \frac{x_n - x_{n-1}}{y_n - y_{n-1}}.100000 \quad lt/100\,km$$

We assumed the distance measurement is in terms of meter instead of km and the variable is being displayed in terms of liter over 100 km. That is why we multiplied the result by 100000.

According to table 2, x_1 and x_0 are D4 and D3 cells respectively and y_1 and y_0 are C4 and C3 cells respectively, the following formula must be written in C3 cell in order to get the instant consumption at the first second in terms of "lt/100 km".

=(D3-D4)/(C4-C3)*100000

The average fuel consumption (the cell G4)

The only difference between average consumption and instant consumption is the necessity that while we try to choose the distance between two points as short as possible for instant consumption, for average consumption it is enough to choose distance from starting point to the point we are on. According to this, we get average consumption value as below.

$$\text{Average Consumption} = \frac{x_n - x_0}{y_n - y_0}.100000 \quad lt/100km$$

Consequently, the formula we must write in cell G4 must be as follows:

=(D3-D4)/(C4-C3)*100000

In this formula, writing D3 instead of D3 and writing C3 instead of C3, will make these cells to be invariant instead of changing relatively when we copy the formula to subjacent lines.

Distance to be taken with remaining fuel (the cell H4)

As the quantity of the fuel we have is written in the cell D4 and the average quantity of the fuel consumed till that time is written in G4 cell in terms of "lt/km" if the car goes on consuming the fuel like this with a simple ratio, the distance that can be taken may be calculated by writing it in H4 cell with the following formula.

=D4/G4*100

Calculating this value also made students aware of the meaning of the calculation of distance to be taken with remaining fuel such that this value means that the distance if the car continues to proceed with the same conditions.

The average speed (the cell I4)

The calculation of the average speed can be made with the ratio of the sum of the distance taken to the total time the car took. Since this values are written in the cells C4 and A4 respectively, average speed can be calculated in terms of "km/h" if we write the following formula in cell I4.

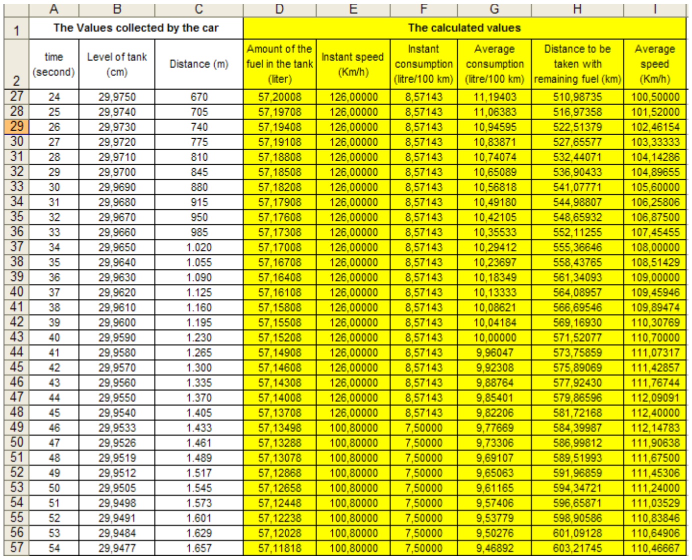
=C4/A4*36/10

All over the electronic table

After copying the formulas we get above for every cell in the 4th row to the subjacent cells, the values can be seen calculated for every second. We should remember the time data in the A column is a natural independent variable. The data in B and C column that are produced by the car according to the real context of it are artificially written by the students with the aim of testing.

This point was another issue that was hard to understand for the students. That is arbitrarily writing the level and the distance was seen as making the all previous effort unessential. On the other hand, the first three data that are given on the white background of the electronic table in table 3 are the data which can be getting after measuring with the help of various receivers or sensors by the car. The data on the colored background are calculated mathematically again by a central chip that is placed on the computer of the car. Here, it is just created a model that shows the computed values. The received values, which of course may be changed by the driving conditions, are being written by the users artificially.

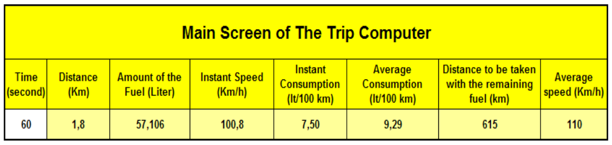
TABLE III. THE LAST VERSION OF ELECTRONIC SPREADSHEET THAT ARTIFICIALLY COMPLETED



	A	B	C	D	E	F	G	H	I
1	The Values collected by the car			The calculated values					
2	time (second)	Level of tank (cm)	Distance (m)	Amount of the fuel in the tank (liter)	Instant speed (Km/h)	Instant consumption (litre/100 km)	Average consumption (litre/100 km)	Distance to be taken with remaining fuel (km)	Average speed (Km/h)
27	24	29,9750	670	57,20008	126,00000	8,57143	11,19403	510,98735	100,50000
28	25	29,9740	705	57,19708	126,00000	8,57143	11,06383	516,97358	101,52000
29	26	29,9730	740	57,19408	126,00000	8,57143	10,94595	522,51379	102,46154
30	27	29,9720	775	57,19108	126,00000	8,57143	10,83871	527,65577	103,33333
31	28	29,9710	810	57,18808	126,00000	8,57143	10,74074	532,44071	104,14286
32	29	29,9700	845	57,18508	126,00000	8,57143	10,65089	536,90433	104,89655
33	30	29,9690	880	57,18208	126,00000	8,57143	10,56818	541,07771	105,60000
34	31	29,9680	915	57,17908	126,00000	8,57143	10,49180	544,98807	106,25806
35	32	29,9670	950	57,17608	126,00000	8,57143	10,42105	548,65932	106,87500
36	33	29,9660	985	57,17308	126,00000	8,57143	10,35533	552,11255	107,45455
37	34	29,9650	1.020	57,17008	126,00000	8,57143	10,29412	555,36646	108,00000
38	35	29,9640	1.055	57,16708	126,00000	8,57143	10,23697	558,43765	108,51429
39	36	29,9630	1.090	57,16408	126,00000	8,57143	10,18349	561,34093	109,00000
40	37	29,9620	1.125	57,16108	126,00000	8,57143	10,13333	564,08957	109,45946
41	38	29,9610	1.160	57,15808	126,00000	8,57143	10,08621	566,69546	109,89474
42	39	29,9600	1.195	57,15508	126,00000	8,57143	10,04184	569,16930	110,30769
43	40	29,9590	1.230	57,15208	126,00000	8,57143	10,00000	571,52077	110,70000
44	41	29,9580	1.265	57,14908	126,00000	8,57143	9,96047	573,75859	111,07317
45	42	29,9570	1.300	57,14608	126,00000	8,57143	9,92308	575,89069	111,42857
46	43	29,9560	1.335	57,14308	126,00000	8,57143	9,88764	577,92430	111,76744
47	44	29,9550	1.370	57,14008	126,00000	8,57143	9,85401	579,86596	112,09091
48	45	29,9540	1.405	57,13708	126,00000	8,57143	9,82206	581,72168	112,40000
49	46	29,9533	1.433	57,13498	100,80000	7,50000	9,77669	584,39987	112,14783
50	47	29,9526	1.461	57,13288	100,80000	7,50000	9,73306	586,99812	111,90638
51	48	29,9519	1.489	57,13078	100,80000	7,50000	9,69107	589,51993	111,67500
52	49	29,9512	1.517	57,12868	100,80000	7,50000	9,65063	591,96859	111,45306
53	50	29,9505	1.545	57,12658	100,80000	7,50000	9,61165	594,34721	111,24000
54	51	29,9498	1.573	57,12448	100,80000	7,50000	9,57406	596,65871	111,03529
55	52	29,9491	1.601	57,12238	100,80000	7,50000	9,53779	598,90586	110,83846
56	53	29,9484	1.629	57,12028	100,80000	7,50000	9,50276	601,09128	110,64906
57	54	29,9477	1.657	57,11818	100,80000	7,50000	9,46892	603,21745	110,46667

Table 3 can be depicted as the medium in which the calculations are done. Certainly the car does not present this data as it is in table 3. With the help of a different interface the data in table 3 can be displayed on the screen as the time passes. In table 4, the electronic table that displays TC data on screen is shown in the case of the change of the time variable by the user.

TABLE IV. THE MAIN SCREEN OF THE TRIP COMPUTER



Main Screen of The Trip Computer							
Time (second)	Distance (Km)	Amount of the Fuel (Liter)	Instant Speed (Km/h)	Instant Consumption (lt/100 km)	Average Consumption (lt/100 km)	Distance to be taken with the remaining fuel (km)	Average speed (Km/h)
60	1,8	57,106	100,8	7,50	9,29	615	110

In the main screen, the behavior of the trip computer can be simulated by changing the cell which the time is written in.

Writing a specific time in term of second simulate the time that car is being driven. When the spreadsheet user enters a certain time the trip computer values are being displayed. By this way, the model provides the opportunity of observing the behavior of the values by travelling in the time.

E. Discussion on the reflections

At the last stage of the modeling work, a discussion session conducted to reach theoretical and practical results.

Instructor: *thanks to the model, which you created, it is understood again that how much mathematics is in our life and most of the technological innovations are products of mathematical concepts. Actually, it is understood that TC produces useful data for us simply by applying a derivative operation. You can reach the instant fuel consumption by taking the limit of average fuel consumption function between two points by closing up the distance between two points to zero. But the method advised in your model finds it enough to choose lower values rather than closing up the distance between two points to zero. Are the calculated fuel consumption values really true?*

Students-C: *No! The consumption value that our model calculated, is not totally true, just an approximation.*

Instructor: *You are right! But, you should not understand the word "approximation" as the TC values is just an approximation.*

Student-A: *Why? It is really seen as an approximation.*

Instructor: *Mathematically yes! The mathematics is devoted to find perfect results. But sometimes this is not completely possible. In the case of infinitesimal calculus, the concept of derivative, which is a special limit operation, calculates the precise result when the concept is used with its complete formal version. If you use the informal and premature version of the concept as above, of course the result is obtained approximately. This approximation is just in the sense of formal mathematics. In the real life, you cannot make the differential operator really zero. The more you chose the differential operator close to zero the more exactly you can calculate the result. On the other hand, in the real sense you do not need the exact result usually. Did you see a car that shows the speed is 92.885 or instant fuel consumption is 5.9562?*

Student-B: *I see! You mean that the value that the car display is enough precise.*

Student-C: *Mathematics gives us a chance of calculating the precise result whatever we need. In this case, we found the exact result that is sufficient for this context.*

It can be observed that this modeling task also enables students to understand the comparison of symbolic exact computation and approximate computation. So, the role mathematics in the real world has been also awaked.

Furthermore, students also found the opportunity of understanding some behavior of the trip computer which is mysterious for some mathematically illiterate people. When the distance and level of the tank values filled until the tank becomes completely empty, a strange thing has been observed at the first look as in a real car's trip computer. The distance to be taken with the remaining fuel was being decreased and increased surprisingly. Students realized that when the values have been setup as lower instant fuel consumption the distance to be taken with the remaining fuel is increased. Knowing the background of calculations helped students to understand this issue.

IV. Result and Discussion

We may not get the result that the calculation method that the car firms use for a navigation computer of a real car is not one to one the same as the method analyzed in this study. However, the artificial navigation computer produced at the end of the study gives a basic idea about the working principle of a real navigation computer.

Additionally, in this study it is illustrated that the volume of the car fuel tank can be calculated with the help of integral concept. In many cars the fuel quantity can be prosecuted only in the category of level. With the system called navigation computer also the fuel quantity of the car can be prosecuted by adding a calculation module like the example which is described above. Certainly, a figure that belongs to a real fuel tank may not be conveyed easily with the help of bivariate functions as it is designed in this study. But it is possible to convey every type of three dimension figure with the help of a specific computer program.

In the context of modeling, this study also showed an example of how the real world can be understood from a mathematical model. At this point, it can be said that any modeling work, like an algebraic equation of a word problem, can help to understand the real situation in the problem. On the other hand, the question is how a computer assisted mathematical model helps to understand the real situation? We can answer this question with the view of expressive modeling [1].

In this modeling work, the main factor of understanding the behavior of trip computer was construction of the mathematical structure as it is explained in the sense of expressive modeling. Using a simple table processor has been facilitated to setup the computer from mathematical language as Grossmann, Masalski and Ozmen also agree [3, 4, 5].

At the last, the method developed in this study can be also considered as an application in which the concepts as derivative, differential and integral are used as integrated. In this respect, it can be said that as Kabaca and Mirasyedioglu, Dede and Argun suggest the mathematical concepts can be concretized with the help of MS Excel electronic spreadsheet [5, 6].

References

[1] Doerr, H.M. and Pratt, D. (2008), The Learning of Mathematics and Mathematical modeling, In Research on Technology in the Teaching and Learning of Mathematics Volume I: Research Syntheses edited by M. K. Heid, G. W. Blume (pp. 259-285) Information Age Publishing.

[2] Masalski, W. (1999). How to use to the spreadsheet as a tool in the secondary school mathematics classroom, Second Edition. National Council of Teachers of Mathematics Inc. 1906 Association Drive, Reston, Virginia VA 20191-1593.

[3] Ozmen, G. (2004). Elektronik tablolar ile kısmi diferansiyel denklemlerin çözümü. İMÖ Teknik Dergi, 3235-3248.

[4] Grossmann, T.A. (1999), Teachers' Forum: Spreadsheet Modeling and Simulation Improves Understanding of Queues, Interfaces, Vol.29, No.3, 88-103

[5] Dede, Y. ve Argun, Z. (2003). Matematik öğretiminde elektronik tabloların kullanımı, Pamukkale Universitesi Egitim Fakültesi Dergisi, 2(14), 113-131.

[6] Kabaca, T. ve Mirasyedioglu, Ş., (2008) A Proposal For Improving The Perception of Differential Concept By Using a Well-Known Table Processor: MS Excel, Korea Society of Mathematical Education, Research in Mathematical Education Vol. 12, No. 3, (193–199).

Design & Analysis of Optical Lenses by using 2D Photonic Crystals for Sub-wavelength Focusing

Rajib Ahmed*[1]
Masters on Photonic Networks Engineering,
Scuola Superiore Sant'Anna,
Pisa, Italy
School of CSE, University of Information Technology & Science,
Dhaka, Bangladesh

Mahidul Haque Prodhan[2]
Dept. of Applied Physics, Electronics & Communication Engineering,
University of Dhaka,
Dhaka, Bangladesh

Rifat Ahmmed[3]
Dept. of Electronics & Telecommunication Engineering,
Rajshahi University of Engineering & Technology (RUET),
Rajshahi, Bangladesh

***Abstract*—2D Photonic lenses (Convex-Convex, Convex-Plane, Plane-Convex, Concave-Concave, Concave-plane, and Plane-Concave) have been designed, simulated and optimized for optical communication using FDTD method. The effect of Crystal structures (Rectangular, Hexagonal, Face centered Cubic (FCC), Body centered Cubic (BCC), variation lattice constant (Λ), hole radius(r), reflective index (n), is demonstrated to get optimized parameters. Finally, with optimized parameters the effect of variation of lens radius on focal lengths and Electrical Field Intensity (Ey) is analyzed. Like optical lens, the focal length of photonic lens is also increased with lens radii, has dependency on optical axis. Moreover, with optimized parameters, Concave-Concave lens have been found as an optimal photonic lens that show sub-wavelength focusing with spatial resolutions-9.22439μm (Rectangular crystal), 7.379512μm (Hexagonal Crystal), 7.840732μm (FCC, BCC).**

Keywords-Photonic crystals; photonic lens; body centered cubic; face centered cubic.

I. INTRODUCTION

In the recent years, a lot of research has been focused on developing micro and nano photonic devices by using Photonic Crystals (PCs). Photonic Crystals (PCs) are artificial structure in which the periodic variation of dielectric constant is used to control the flow of light. PCs may be 1D, 2D & 3D crystal structure. Unlike electronic crystals (which are natural structure) that control the flow of electron by periodic variation of electron potential and follow Schrödinger equation, PCs follow Maxwell's equation [1].

Now a day, from the idea of controlling light by means of Photonic Crystal (PC) has led to many proposals and implementations for novel devices including different types of focusing elements and nano-scale imaging field. This is possible only for negative refraction effect. The traditional limitation on Optical lenses is no lens can focus light onto an area smaller than a square wavelength, overcome by Photonic lenses as Super lenses [2].

We have proposed different photonic lenses (Convex-Convex, Convex-Plane, Plane-Convex, Concave-Concave, Concave-plane, Plane-Concave) with Rectangular, Hexagonal, Face-centered Cubic (FCC), Body centered Cubic (BCC) photonic crystals for sub-wavelength focusing and try to optimize the parameters (Hole Radius(r), Positions, Reflective Index (n), Input Signal wavelength (λ), lattice constant (Λ), Distances or position of lens, Focal length (f), Spatial resolution(Δx)) to get optimal crystal structure (Rectangular) with optimal photonic lens (Concave-Concave). Moreover, focusing of the light is found smaller then square input wavelength. Thus Sub-wavelength focusing has been found for photonic lenses.

The Sub-wavelength focusing has great Effect in nano-scale imaging, coupling large bandwidth waveguide to lower bandwidth waveguide applications and so on.

II. THEORITICAL BACKGROUND

Propagation of light in photonic crystals is described by Maxwell's equations. Solving the equation under the condition of no free charges or currents, we get the following equation

$$\nabla \times \left(\frac{1}{\varepsilon(\mathrm{r})} \nabla \times \mathrm{H}(\mathrm{r})\right) = \left(\frac{\omega}{c}\right)^2 \mathrm{H}(\mathrm{r}) \qquad (1)$$

Equation (1) is known as the master equation [1]. Here, ε (r), H(r), ω and c represent the dielectric constant, the magnetic field distribution, the frequency and the speed of light in vacuum, respectively. Eq. (1) follows the Bloch-Floquet theorem, which proves that waves in 3D periodic media can propagate without scattering. Their behavior governed by a periodic function multiplied by a plane wave. The Bloch modes have the form

$$\boldsymbol{H(r)} = e^{i(k.r-\omega t)} \boldsymbol{H_{n,k}} \qquad (2)$$

With eigenvalues $\omega_n(k)$, where $H_{n,k}$ is a periodic envelope function satisfying:

$$(\nabla + i\boldsymbol{k}) \times \frac{1}{\varepsilon} (\nabla + i\boldsymbol{k}) \times \boldsymbol{H_{n,k}} = \left(\frac{\omega_n(\boldsymbol{k})}{c}\right)^{\mathbf{2}} \boldsymbol{H_{n,k}} \qquad (3)$$

yielding a different Hermitian eigen problem over the primitive cell of the lattice at each Bloch wave vector k.This primitive cell is a finite domain if the structure is periodic in all directions, leading to discrete eigenvalue labeled by n=1, 2… These eigenvalues when plotted with respect to the wave vector, k forms the band diagram or dispersion diagrams [4].

These eigenvalues are periodically repeating functions of k. The values at k are same as the values obtained at k+Gj.

Here, Gj denotes the primitive reciprocal lattice vector [1]. Thus, eigenvalues can be computed for only the wave vector, k. This unit cell repeats to form the entire lattice structure [1].

III. Layout Design And Simulation

We take SiO_2 as a wafer material (reflective index 1.447) with dimension- length-30μm and width-48μm, Gaussian modulated continuous wave as an input signal with wavelength 1.55μm. Both transverse and injection type of the input is Modal. The input signal (Fig. a) used in the simulation expressed as:

$$E_y^{inc}(x, z_{inc})= AT(t)\, F(x, z_{inc}) \sin (wt + \theta_i)(4)$$

Where, A is the field amplitude and F (x, z_{inc}) is the transverse field location at the incident field location z_{inc}. The initial offset θ_i is the phase difference between points in the incidence plane. This offset can be adjusted to define the direction of the incident field.

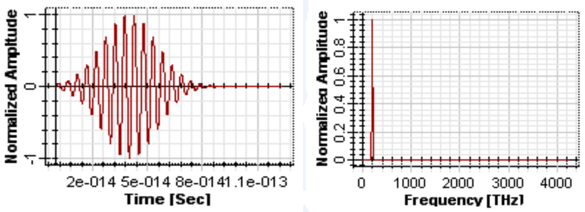
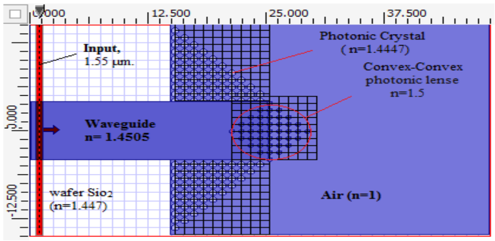
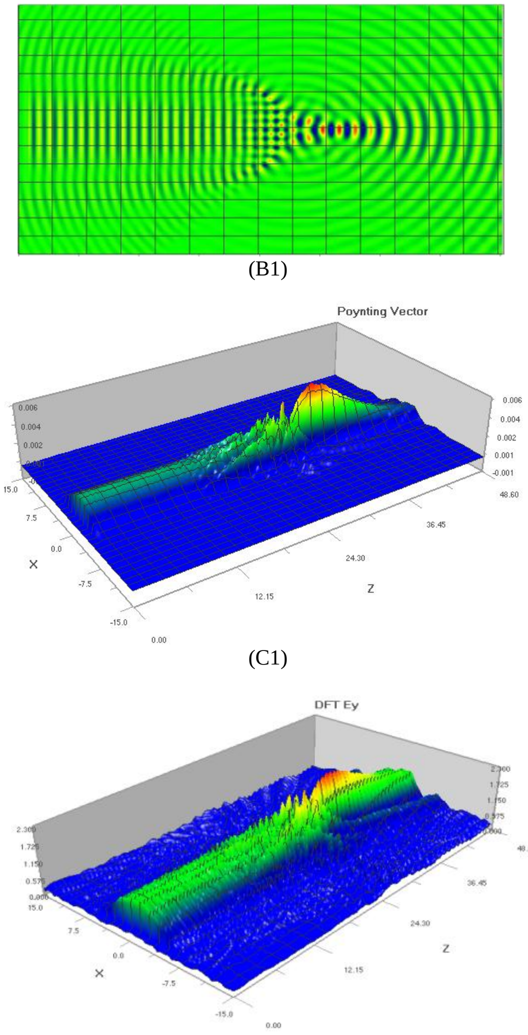


Figure a: Gaussian modulated continuous signal with time and frequency

The propagation of light in the waveguides is simulated by 2D Finite-difference time-domain (FDTD) method using a standard simulator. FDTD is a time-domain numerical method used for modeling the propagation of electromagnetic waves in optical media, which is based on the discretization of Maxwell's equations in differential form in free space. Time domain methods have been found to be very accurate in simulating the propagation dynamics of signals in periodic dielectric media [3].We simulate for TE mode, actual Mesh used- 0.075μm (delta X) X 0.075μm (delta Z) with number of mesh cells 400 (X) and 650 (Z). For result finalization the simulation has been done for 4000 time steps. The Anisotropic Perfectly Matched Layer (APML) boundary condition is used with 20 layers. Figure1 shows the Simulation setup for Convex-Convex lenses, 2D-light propagation during simulation and finally resulted 3D-Poynting Vector & DFT E $_y$ curve. Moreover, Fig.2 and Fig. 3 show the Simulation setup and 2D focusing for Convex-Plane (A2), Plane–Convex (B2), Concave-Concave (C2), Concave-plane (D2), Plane-Concave (E2) Lenses.



(A1)

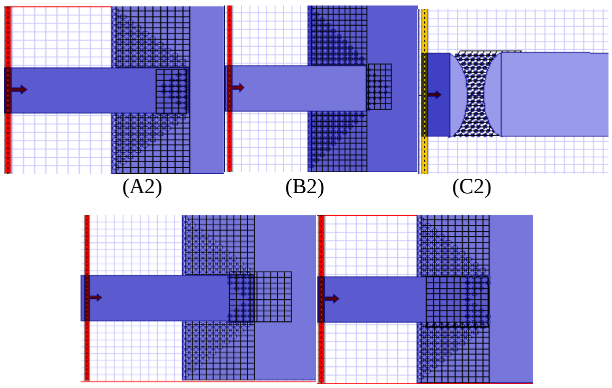


(B1)

(C1)

(D1)

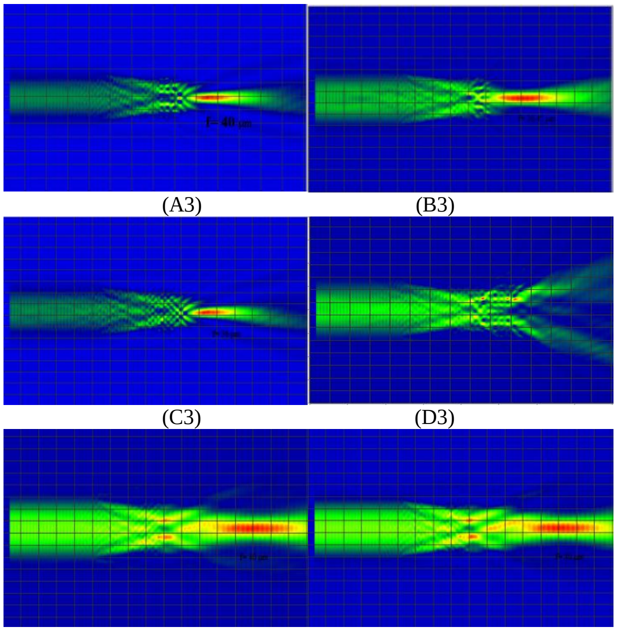
Figure 1. (A1) Experimental Setup, (B1) 2D Simulation, (C1) Poynting Vector, (D1) DFT E y



(A2) (B2) (C2)

(D2) (E2)

Figure 2. Experimental Setup Convex-Plane (A2), Plane–Convex (B2), Concave-Concave (C2), Concave-plane (D2), Plane-Concave (E2) Lenses.



(A3) (B3)

(C3) (D3)

(E3) (F3)

Figure 3. 2D focusing for (A3) Convex-Plane, (B3) Plane Convex, (C3) Concave-Concave, (D3) Concave-plane and (E3) Plane-Concave Lenses.

IV. Result Analysis

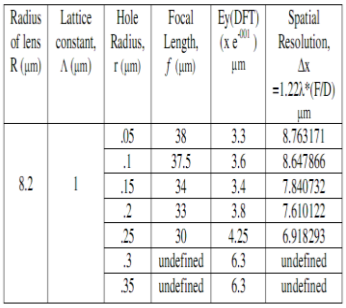
We have done FDTD simulation for implementation of optical focusing with Rectangular, Hexagonal, Face-centered Cubic (FCC), and Body centered Cubic (BCC) photonic crystals. For Convex-Convex lens (FCC), with hole radius(r) increases, the focal length (*f*) and the spatial resolution (Δx) decrease but DFT E_y increases. After, hole radius, r= 0.3μm, we get no fixed focal length & spatial resolution (Table-1). In all case, the lens radius(R=8.2μm) & lattice constant (Λ=1μm) are kept constant.

Similarly, for Convex-Plane, Plane–Convex, Concave-Concave, Concave-plane, Plane-Concave lenses with Rectangular, Hexagonal, Face-centered Cubic (FCC), Body centered Cubic (BCC) photonic crystals the hole radius, distances or position of lenses from waveguide are varied (Table- 1 to Table-10). The optimal values of distances or position of lenses is zero μm, hole radiuses, r= 0.2μm (FCC), r=0.3 μm (BCC, Rectangular), r=0.4 μm (Hexagonal). Moreover, like convention optical lenses, the focal length (*f*) of photonic lenses also decreases with increase of diameter of lens (Table-11). From the variation of reflective index (n) of the lens, the optimal value is found n= 1.4505(Table-12). Finally, the dependence of variation of focal length(*f*), spatial resolution (Δx) on input signal wave length(λ) is analyzed to get optimal values, λ=1.55 μm (for TM Mode)(Table-14) .

But with TE Mode we get no focal length(*f*), spatial resolution (Δx) before λ=1.35μm(Table-13).With optimized parameters (Hole Radius(r), Positions, Reflective Index(n), Input Signal wavelength(λ), lattice constant(Λ) etc. the optimized values of focal lengths(*f*): 40μm, 32μm,34μm,34μm and spatial resolution 9.22μm, 7.37μm, 7.84μm, 7.84μm (for Rectangular, Hexagonal, BCC, FCC respectively) are found. Thus, the best result is found for Rectangular photonic crystal configuration (Table-15).

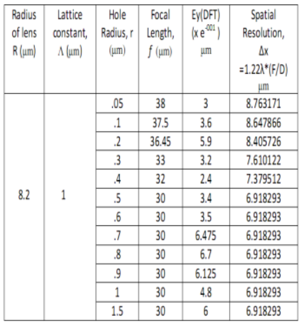
Finally, with optimized parameter and photonic crystal structure (rectangular)- DFT E_y curve (Figure4) for Convex-Plane, Plane–Convex, Concave-Concave, Concave-plane, Plane-Concave Lenses are found. From those DFT E_y curves, the maximum value and concentrations of DFT Ey across zero position (along X-axis) is occurring for Concave-Concave photonic lens. Moreover, sub-wavelength focusing are occurred for all proposed photonic lenses i.e focusing from the lenses will occur less than the half of the input signal wavelength (Fig. 3). Therefore, the limitation of normal glass lenses can be overcome by photonic lenses. Like convention optical lens, photonic lens also having dependence on optical axis and Concave-Concave photonic lens is the optimal configuration for light focusing (Table-15).

Table 1: The variation of hole radius for Convex-Convex lens (FCC Crystal)



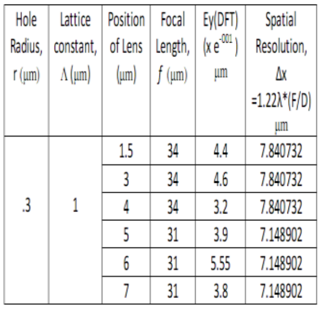
Radius of lens R (μm)	Lattice constant, Λ (μm)	Hole Radius, r (μm)	Focal Length, *f* (μm)	Ey(DFT) (x e^{-001}) μm	Spatial Resolution, Δx =1.22λ*(F/D) μm
8.2	1	.05	38	3.3	8.763171
		.1	37.5	3.6	8.647866
		.15	34	3.4	7.840732
		.2	33	3.8	7.610122
		.25	30	4.25	6.918293
		.3	undefined	6.3	undefined
		.35	undefined	6.3	undefined

Table 2: The variation of hole radius for Plane-Convex lens (Rectangular Crystal)



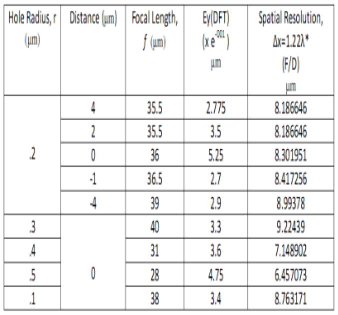
Radius of lens R (μm)	Lattice constant, Λ (μm)	Hole Radius, r (μm)	Focal Length, *f* (μm)	Ey(DFT) (x e^{-001}) μm	Spatial Resolution, Δx =1.22λ*(F/D) μm
8.2	1	.05	38	3	8.763171
		.1	37.5	3.6	8.647866
		.2	36.45	5.9	8.405726
		.3	33	3.2	7.610122
		.4	32	2.4	7.379512
		.5	30	3.4	6.918293
		.6	30	3.5	6.918293
		.7	30	6.475	6.918293
		.8	30	6.7	6.918293
		.9	30	6.125	6.918293
		1	30	4.8	6.918293
		1.5	30	6	6.918293

Table 3: The variation of Position of Plane-Convex lens



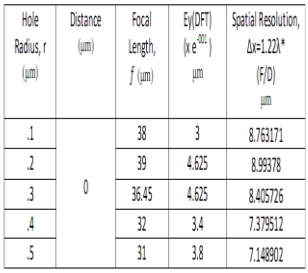
Hole Radius, r (μm)	Lattice constant, Λ (μm)	Position of Lens (μm)	Focal Length, *f* (μm)	Ey(DFT) (x e^{-001}) μm	Spatial Resolution, Δx =1.22λ*(F/D) μm
.3	1	1.5	34	4.4	7.840732
		3	34	4.6	7.840732
		4	34	3.2	7.840732
		5	31	3.9	7.148902
		6	31	5.55	7.148902
		7	31	3.8	7.148902

Table 4: The variation of Hole Radius & Position of Convex -Convex lens (Rectangular Crystal)



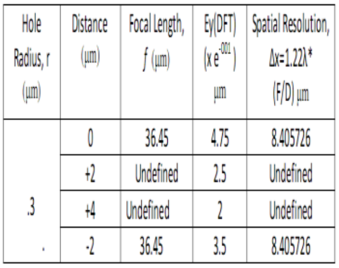
Hole Radius, r (μm)	Distance (μm)	Focal Length, *f* (μm)	Ey(DFT) (x e^{-001}) μm	Spatial Resolution, Δx=1.22λ* (F/D) μm
.2	4	35.5	2.775	8.186646
	2	35.5	3.5	8.186646
	0	36	5.25	8.301951
	-1	36.5	2.7	8.417256
	-4	39	2.9	8.99378
.3	0	40	3.3	9.22439
.4		31	3.6	7.148902
.5		28	4.75	6.457073
.1		38	3.4	8.763171

Table 5: The variation of Hole Radius of Convex -Convex lens (Hexagonal Crystal)



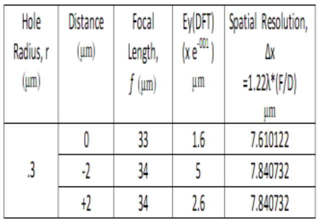
Hole Radius, r (μm)	Distance (μm)	Focal Length, *f* (μm)	Ey(DFT) (x e^{-001}) μm	Spatial Resolution, Δx=1.22λ* (F/D) μm
.1	0	38	3	8.763171
.2		39	4.625	8.99378
.3		36.45	4.625	8.405726
.4		32	3.4	7.379512
.5		31	3.8	7.148902

Table 6: The variation of Position of Convex-Convex lens (Hexagonal Crystal)



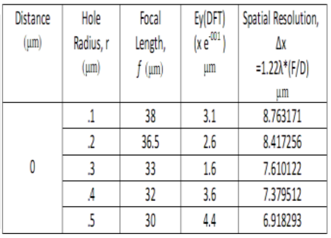
Hole Radius, r (μm)	Distance (μm)	Focal Length, *f* (μm)	Ey(DFT) (x e^{-001}) μm	Spatial Resolution, Δx=1.22λ* (F/D) μm
.3	0	36.45	4.75	8.405726
	+2	Undefined	2.5	Undefined
	+4	Undefined	2	Undefined
.	-2	36.45	3.5	8.405726

Table 7: The variation of Position of Convex–Convex lens (BCC Crystal)



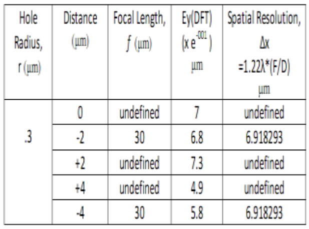
Hole Radius, r (µm)	Distance (µm)	Focal Length, f (µm)	Ey(DFT) (x e^{-001}) µm	Spatial Resolution, Δx =1.22λ*(F/D) µm
.3	0	33	1.6	7.610122
	-2	34	5	7.840732
	+2	34	2.6	7.840732

Table 8: The variation of Hole Radius of Convex–Convex lens (BCC Crystal)



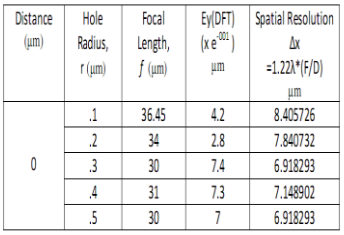
Distance (µm)	Hole Radius, r (µm)	Focal Length, f (µm)	Ey(DFT) (x e^{-001}) µm	Spatial Resolution, Δx =1.22λ*(F/D) µm
0	.1	38	3.1	8.763171
	.2	36.5	2.6	8.417256
	.3	33	1.6	7.610122
	.4	32	3.6	7.379512
	.5	30	4.4	6.918293

Table 9: The variation of Position of Convex-Convex lens (FCC Crystal)



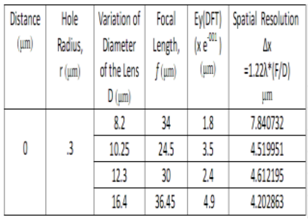
Hole Radius, r (µm)	Distance (µm)	Focal Length, f (µm)	Ey(DFT) (x e^{-001}) µm	Spatial Resolution, Δx =1.22λ*(F/D) µm
.3	0	undefined	7	undefined
	-2	30	6.8	6.918293
	+2	undefined	7.3	undefined
	+4	undefined	4.9	undefined
	-4	30	5.8	6.918293

Table 10: The variation of hole radius of Convex-Convex lens (FCC Crystal)



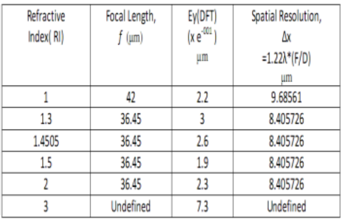
Distance (µm)	Hole Radius, r (µm)	Focal Length, f (µm)	Ey(DFT) (x e^{-001}) µm	Spatial Resolution Δx =1.22λ*(F/D) µm
0	.1	36.45	4.2	8.405726
	.2	34	2.8	7.840732
	.3	30	7.4	6.918293
	.4	31	7.3	7.148902
	.5	30	7	6.918293

Table 11: The variation of Diameter of Lens (Convex-Convex) for Rectangular Crystal



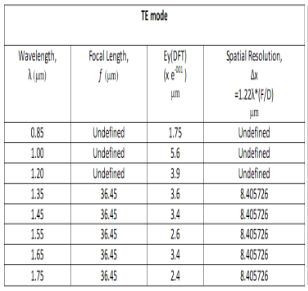
Distance (µm)	Hole Radius, r (µm)	Variation of Diameter of the Lens D (µm)	Focal Length, f (µm)	Ey(DFT) (x e^{-001}) (µm)	Spatial Resolution Δx =1.22λ*(F/D) µm
0	.3	8.2	34	1.8	7.840732
		10.25	24.5	3.5	4.519951
		12.3	30	2.4	4.612195
		16.4	36.45	4.9	4.202863

Table 12: The variation of Refractive Index for Convex-Convex Lens (Rectangular Crystal)



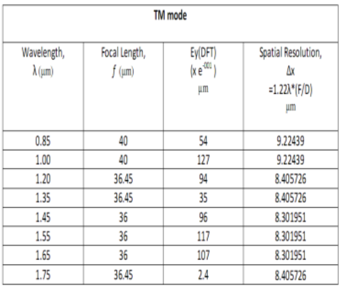
Refractive Index(RI)	Focal Length, f (µm)	Ey(DFT) (x e^{-001}) µm	Spatial Resolution, Δx =1.22λ*(F/D) µm
1	42	2.2	9.68561
1.3	36.45	3	8.405726
1.4505	36.45	2.6	8.405726
1.5	36.45	1.9	8.405726
2	36.45	2.3	8.405726
3	Undefined	7.3	Undefined

Table 13: The variation of wavelength for Convex- Convex Lens (Rectangular Crystal) with TE mode



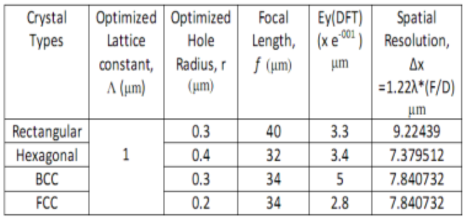
TE mode			
Wavelength, λ (µm)	Focal Length, f (µm)	Ey(DFT) (x e^{-001}) µm	Spatial Resolution, Δx =1.22λ*(F/D) µm
0.85	Undefined	1.75	Undefined
1.00	Undefined	5.6	Undefined
1.20	Undefined	3.9	Undefined
1.35	36.45	3.6	8.405726
1.45	36.45	3.4	8.405726
1.55	36.45	2.6	8.405726
1.65	36.45	3.4	8.405726
1.75	36.45	2.4	8.405726

Table 14: The variation of wavelength for Convex-Convex Lens (Rectangular Crystal) with TM mode

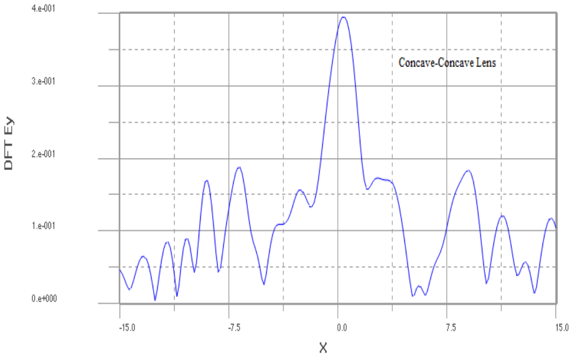


TM mode			
Wavelength, λ (µm)	Focal Length, f (µm)	Ey(DFT) (x e^{-001}) µm	Spatial Resolution, Δx =1.22λ*(F/D) µm
0.85	40	54	9.22439
1.00	40	127	9.22439
1.20	36.45	94	8.405726
1.35	36.45	35	8.405726
1.45	36	96	8.301951
1.55	36	117	8.301951
1.65	36	107	8.301951
1.75	36.45	2.4	8.405726

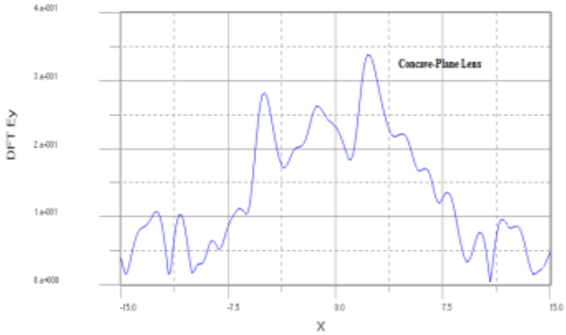
Table 15: Optimization of parameters for crystals (Rectangular Crystal)



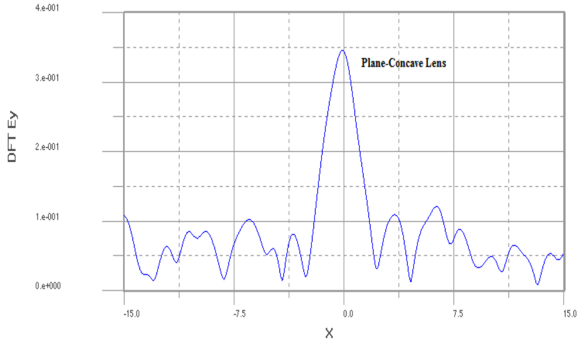
Crystal Types	Optimized Lattice constant, Λ (µm)	Optimized Hole Radius, r (µm)	Focal Length, f (µm)	Ey(DFT) (x e^{-001}) µm	Spatial Resolution, Δx =1.22λ*(F/D) µm
Rectangular	1	0.3	40	3.3	9.22439
Hexagonal		0.4	32	3.4	7.379512
BCC		0.3	34	5	7.840732
FCC		0.2	34	2.8	7.840732



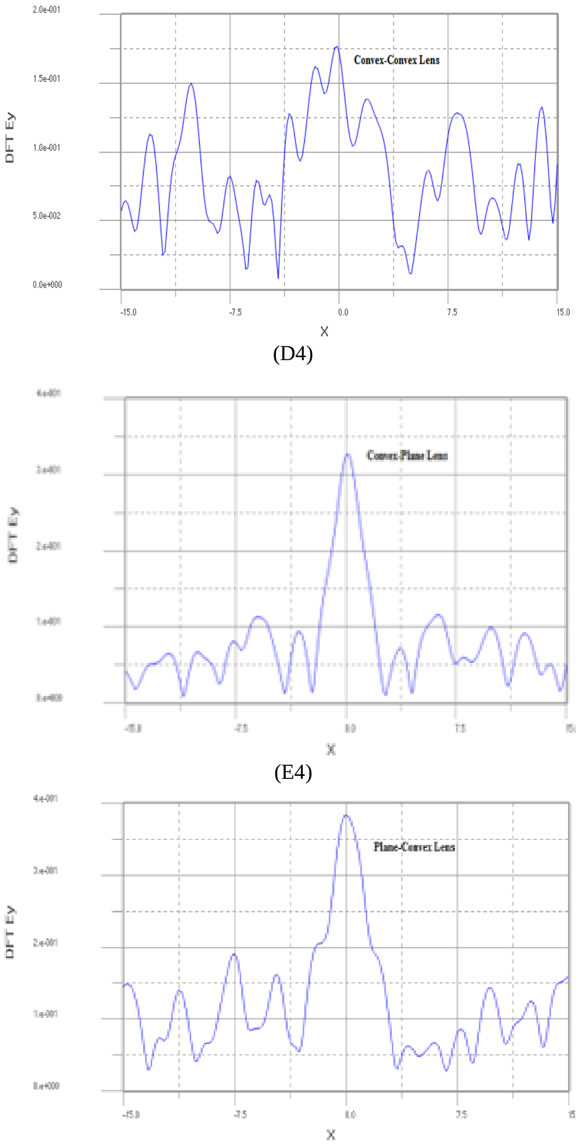
(A4)



(B4)



(C4)



(D4)

(E4)

(F4)

Figure 4. DFT Ey curves for Concave-Concave(A4), Concave-Plane(B4), Plane-Concave(C4), Convex-Convex (D4), Convex-Plane(E4), Plane-Covex Lens(F4)

V. Conclusion

An ideal case is considered without any temperature effect but temperature will cause the change of the reflective index and consequently to focal length (*f*) and the spatial resolution (Δx) for TM Mode. For TE Mode, it will remain unaffected. Moreover, unlike optical lenses- Sub-wavelength focusing has been found for photonic lenses but axis dependence and change of the focal length (*f*) with diameter is almost same. With optimize parameters: Hole radiuses (r), 0.2µm (FCC), 0.3 µm (BCC, Rectangular), 0.4 µm (Hexagonal), lattice constant, Λ =1, distances or position of lenses is zero µm, reflective index n=1.4505, focal length(*f*): 40µm, 32µm, 34µm, 34µm and spatial resolution(Δx): 9.22µm, 7.37µm, 7.84µm, 7.84µm (for Rectangular, Hexagonal, BCC, FCC). Finally, with optimized parameters, Concave-Concave photonic lens is found as optimal photonic lens for Rectangular crystal structure that shows.

Acknowledgement

We would like to thanks all reviewers for their valuable comments. All thanks due especially to my parents for their prayer and inspiration, which has helped me in becoming what I am today.

References

[1] J. D. Joannopoulos, R. D. Meade, and J. N. Winn “Photonic Crystals: Molding the Flow of Light”, Princeton University Press (1999).

[2] Qi Wu, John M. Gibbons and Wounjhang Park, “Graded Negative Index Lens by Photonic Crystal”, Optics Express, Vol.16, No. 21, 16942-16949, (2008).

[3] A. Taflove,“Compuatational Electrodynamics: The Finite-Difference Time-Domain Method”, Artech House, Boston, MA (1995).

[4] N. Palka, W. Ciurapinski and J. Wróbel, “Focusing with 2D Square Photonic Crystal with Concavo-Concavo Boundaries”, Optical and Acoustical Methods in Science and Technology, Vol.116, No. 3, 368-370, (2009).

[5] I. V. Minin, O. V. Minin, Y. R. Triandaphilov and V. V. Kotlyar, “Sub-wavelength Diffractive Photonic Crystal Lens”, progress in Electromegneties Research B, Vol. 7, 257-264, (2008).

[6] Hamza Kurt, “Graded index photonic crystals”, Optics Express, Vol.15, No. 3, 1241-1253, (2007).

[7] Zhaolin Lu, Shouyuan Shi, Christopher A. Schuetz, Janusz A. Murakowski, and Dennis W. Prather, “Three-dimensional photonic crystal flat lens by full 3D negative refraction”, Optics Express, Vol.13, No. 15, 5592-5599, (2005).

[8] X. Wang, Z. F. Ren and K. Kempa, “Unrestricted superlensing in a triangular two-dimensional photonic crystal”, Optics Express, Vol.12, No. 13, 2919-2924, (2004).

Authors Profile

Rajib Ahmmed was born in Kishoregonj, Bangladesh on 05 May, 1986. He received the B. Sc. Honours and M.Sc degree in Applied Phsics, Electronics and Communication Engineering from Dhaka University, Dhaka, Bangladesh, in 2008. He is a Lecturer (study leave) of University of Information Technology & Sciences (UITS), Baridhara, Dhaka, Bangladesh. Now, he doing Erasmus Mundus MASters on Photonic NETworks engineering (MAPNET), Institute of Scuola Superiore Sant’ Anna, Pisa, Italy and Phone: +393892320071. His research interest concerns on the photonic related works.

Mahidul Haque Prodhan is with the Applied physics, Electronics & Communication Engineering, Institute of University of Dhaka, Dhaka-1000, Bangladesh;

Rifat Ahmmed is with the Electronics and Telecommunication Engineering department, Institute of Rajshahi University of engineering and Technology, Rajshahi-6204, Bangladesh, Phone: +880-1914961164;

Secure Optical Internet: A Novel Attack Prevention Mechanism for an OBS node in TCP/OBS Networks

K. Muthuraj
Computer science and Engineering Department
Pondicherry Engineering College
Puducherry, India

N. Sreenath
Computer science and Engineering Department
Pondicherry Engineering College
Puducherry, India

***Abstract*—Optical Internet has become a strong development and its commercial use is growing rapidly. Due to transparency and virtual sharing infrastructure, they provide ultra-fast data rates with the help of optical burst switching technology, which transmits data in the form of bursts. From the security perspective, one of the OBS nodes in the optical network is compromised, causes the vulnerability. This paper is dealt to identify the vulnerabilities and named as burst hijacking attack and provide the prevention mechanism for the same. The NSFnet 14 nodes and the ns2 simulator with modified nOBS patch is used to simulate and verify the security parameters.**

Keywords-optical internet security; burst hijacking attack; threats and vulnerabilities in TCP/OBS networks.

I. Introduction

The benefits of optical internet have been known for quite awhile; but it was not until the invention of wavelength division multiplexing (WDM) that the potential of fiber was fully realized [1]. This divides the available bandwidth of the fiber into a number of separate wavelength channels and allows tens or hundreds of wavelength channels to be transmitted over a single optical fiber at a rate of 10 Gb/s/channel and beyond. This means that the data rate can reach 10 Tb/s in each individual fiber [2].

To carry IP traffic over WDM networks three switching technologies exist namely optical circuit switching (OCS), optical packet switching (OPS) and optical burst switching (OBS).Optical circuit switching, also known as lambda switching, can only switch at the wavelength level, and is not suitable for bursty internet traffic [3-5]. Optical packet switching, which can switch at the packet level with a fine granularity, is not practical in the foreseeable future. The two main obstacles are lack of random access optical buffers, and optical synchronization of the packet header and payload. Optical burst switching can provide fine granularity than optical circuit switching, and does not encounter the technical obstacles that optical packet switching faces. OBS is considered the most promising form of optical switching technology, which combines the advantages and avoids the shortcomings of OCS and OPS as tabulated in Table1 [6 -8].

OBS can provide a cost effective means of interconnecting heterogeneous networks regardless of lower-level protocols used in optical internet [9-11]. For example, an OBS network is able to transport 10 GB/s Ethernet traffic between two sub-networks without the need to interpret lower level protocols, or to make two geographically distant wireless networks to act as an integrated whole without protocol translations. The illustration of optical burst switching networks in the optical internet as shown in below Fig.1.

TABLE I. SCOPE OF SWITCHING TECHNOLOGY

Technique	Bandwidth	Latency	Buffer	Overhead	Adaptive
OCS	Low	High	-	Low	Low
OPS	High	Low	Yes	High	High
OBS	High	Low	-	Low	High

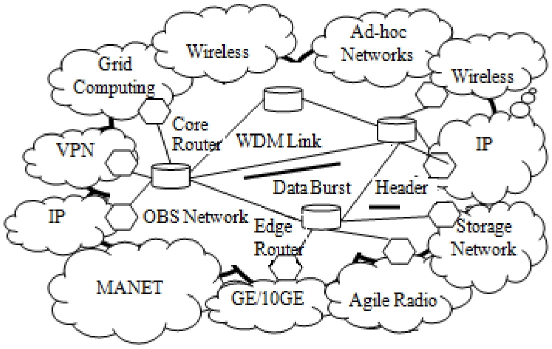


Figure 1. Illustration of optical burst switching network

In OBS networks, there is a strong separation between the control and data planes, which allows for great network manageability and flexibility. In addition, its dynamic nature leads to high network adaptability and scalability, which makes it quite suitable for transmission of bursty traffic. Unfortunately, OBS networks suffer from security vulnerabilities. Since every data burst is pass through the intermediate OBS routers. If one of the OBS intermediate routers is compromised, it causes security issues and denial of services [12-15].

The remainder of this paper is organized as follows. The architecture of OBS and about in-band and out-of-band signaling with its functional diagram is described in Section II. The Section III explains the TCP over OBS networks in Optical Internet. The Section IV demonstrates the main objective of this paper that is the identification of the attack on OBS node in TCP/OBS networks in Optical Internet as named as Burst hijacking attack. Section V depicts the attack prevention mechanism for the same. The simulation results are shown in section VI. Finally we conclude and notify the future work in Section VII.

II. Optical Burst Switching Architecture

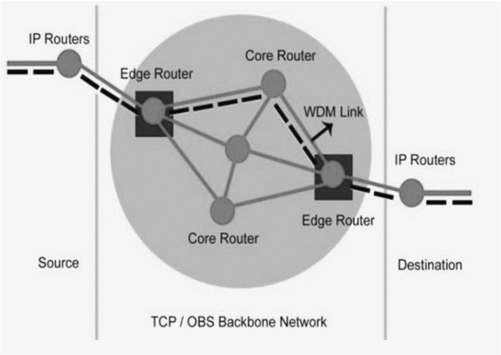


Figure 2. Architecture of optical burst switching

The pictorial representation of OBS architecture is shown in the above Fig. 2. In general, OBS network is composed of two types of routers, namely edge routers and core routers. Edge routers represent the electronic transit point between the burst-switched backbone and IP routers in an Optical Internet. The assembling of bursts from IP packets and disassembling of burst into IP packets is carried out at these edge routers. Core routers are connected to either edge routers or core routers. It transfers the incoming optical data into an outgoing link in the optical form without conversion of electronic form. In OBS, the basic switching entity is burst which contains the number of encapsulated packets. For every burst, there is a corresponding Burst Control Header (BCH) to establish a path from source to destination [16 -18]. BCH of a connection is sent prior to the transmission of Data Burst (DB) with specific offset time on the same wavelength channel is termed as In – band signaling shown in Fig. 3.

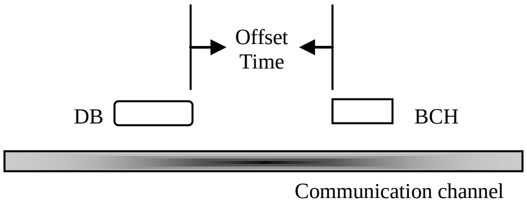


Figure 3. In – band signaling

All BCH's of various connections are sent on the same control channel and their corresponding DBs will sent on the different channels with specific offset time named as out – of – band signaling is shown in Fig. 4.

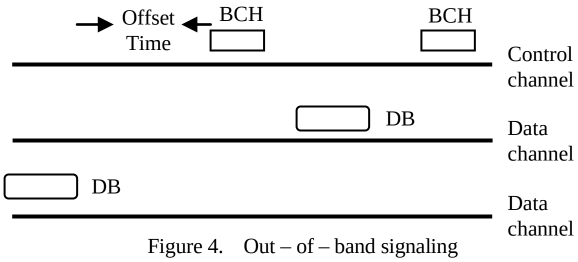


Figure 4. Out – of – band signaling

The Offset time is the transmission time gap between the BCH and DB, which is used to allow the control part in intermediate core nodes to reserve the required resources for the onward transmission of bursts.

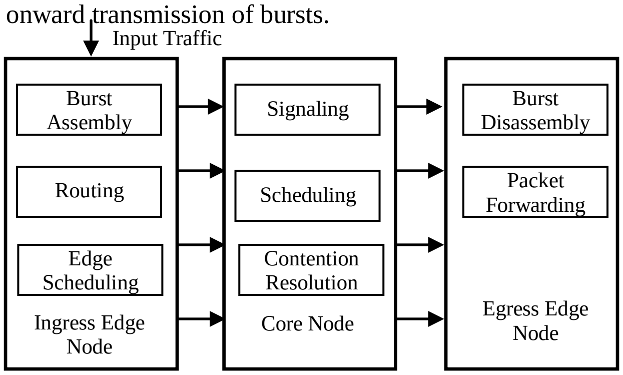


Figure 5. OBS functional diagram

The OBS functional diagram is shown in Fig. 5. It describes the ingress node is responsible for burst assembly, routing, wavelength assignment and scheduling of burst at the edge node. The core node is responsible for signaling and contention resolution. The egress edge node is responsible for disassembling the burst and forwarding the packets to the higher network layer [19-21].

III. TCP over OBS Network

In a TCP/IP network, IP layer is involved in routing of packets, congestion control and addressing the nodes. When OBS is introduced in the network, it takes care of routing of data and congestion control. The routing information computed by IP layer need not be considered by OBS routers. It is because, the routes at the OBS are computed based on number of hops and wavelength availability. However, the addressing of the various nodes in the network is not taken care by OBS by default. Hence the functionality of IP may be limited to addressing and packet formation. Due to above reasons, this proposal consider the stack TCP/OBS rather than TCP/IP/OBS. This is shown in Fig. 6.

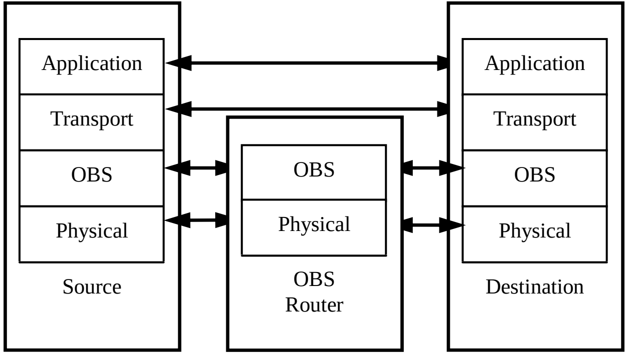


Figure 6. TCP/OBS Layer Architecture

IV. Burst Hijacking Attack

In Optical Internet, an optical virtual source node is inevitable for multicast routing, which is an optical node holds both wavelength splitting capability and wavelength conversion capability. It can transmit an incoming burst to multiple destinations on any wavelength. If it is compromised, a new type of attack is possible named as burst hijacking attack. During the data transmission, Burst Control Header (BCH) is converted from an optical form to electronic form and is processed at every intermediate core node. The core node is to receive the BCH and setup the path for the corresponding Data Burst (DB) and forward to the next intermediate optical node until it reaches the egress node. If a compromised optical virtual source node receives, it can maliciously create a copy of original BCH and modifies its value to setup a path to a malicious destination then the corresponding DB will travel into the original destination as well as the malicious destination as shown in Fig. 7. The malicious destination will not send the acknowledgment for this hijacked burst and it escapes from being caught. Thus it compromises the authentication of data burst and also denial of service. This threat can be detected and named as Burst hijacking attack.

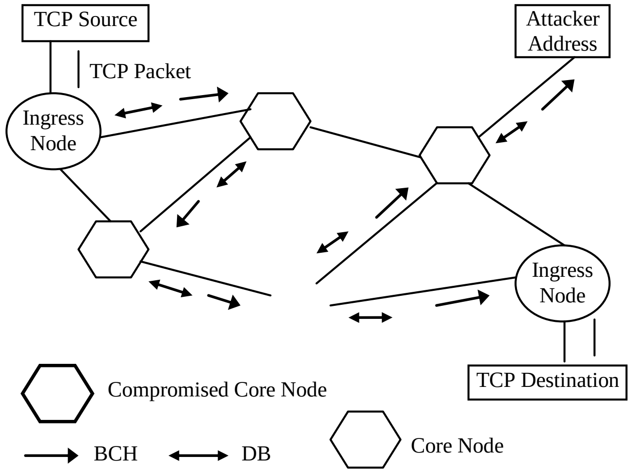


Figure 7. Burst hijacking attack

The provisioning of security has two aspects, attack detection and attack prevention. In attack detection the intermediate optical nodes are being monitored using trusted optical nodes. These trusted optical nodes use the statistical report to identify the malicious nodes. When an intermediate core router gets BCH, it collects the statistical information like burst id, source, destination, number of packets present inside the burst and the size of the burst. If the buffer is not present in the particular intermediate node then the collected statistical information is sent immediately to the trusted optical node. If it has buffers then it stores the statistical information and starts a timer. Once the timer gets expired or the buffer gets full, it sends the statistical information to the trusted optical node. The collected statistical information is stored in the buffer table of the trusted optical node based on the burst id. The statistical information is observed for some predetermined number of seconds and it should be analyzed and determines whether the node is behaving maliciously or not.

In Burst hijacking attack, a new connection established between the compromised intermediate virtual source node and destination node. Just to escape from being caught, the intelligent compromised virtual source node changes the burst id every time and creates a new connection between the intermediate core node and destination. In this case, the trusted optical node should check the burst size, number of packets inside the burst and detects the malicious optical node.

V. Attack Prevention Mechanism

```
recv (struct * PACKET packet)
{
Determine nodeType from packet.
if ((nodeType = 'intermediate core node')
OR
(nodeType = 'egress node'))
{
    a) Extract burst id, source, destination, num_of_packets, burst_size from the packet.
    b) Create a new packet and store the extracted information inside the new packet.
    c) Send the new packet to the trusted node
}
else if (nodeType == 'trusted_node')
{
    a) Extract statistics from packet.
    b) Insert the statistics into the linked list based on burst id.
    c) Collect some more statistics.
    d) Now extract the source, destination and burst id from the linked list head.
    e) For burst hijacking attack, verify a new connection is established in virtual source node and burst_id or statistics matches with the original source and destination.
    f) If the node's trust value reaches below threshold, inform other nodes.
}
}
```

VI. Simulation Resesults

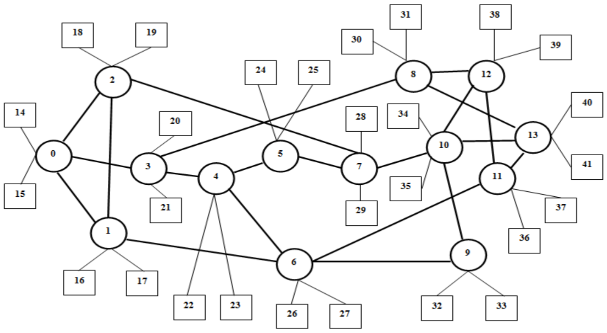


Figure 8. NSFNet topology with nodes 0 to 13

Topology	:	NSFNet
Number of Optical Nodes	:	14

Number of Electronic Nodes	:	28
Number of TCP/IP Connection	:	10
Max. Number of attacker nodes	:	03
Max. Number of packets	:	200
Max Lambda	:	20
Link Speed	:	1GB
Switch Time	:	0.000005

The simulations are done using nOBS, an ns2 based network simulator. NSFNet topology is used to demonstrate the effect of the BCH flooding attack as shown in Fig. 8. Nodes 0 to 13 represent the optical nodes and 14 to 41 represent the electronic nodes. The optical network is modeled with 1Gbps bandwidth and 10ms propagation delay. The TCP/IP links have 155 Mbps bandwidth each with 1 ms link propagation delay. In the beginning let us assume that there are no compromised nodes in the network. In that case, number of bursts sent by the ingress edge node is almost equal to number of bursts received by the egress edge node as shown in Fig. 9. Indicates that the number of bursts that are hijacked at that particular interval of time.

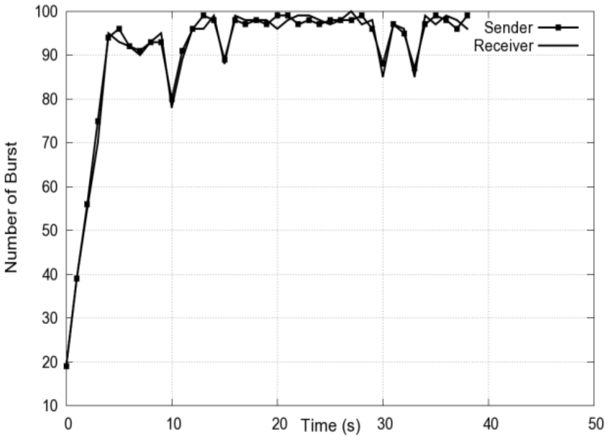


Figure 9. Number of bursts sent/received without any attacker nodes.

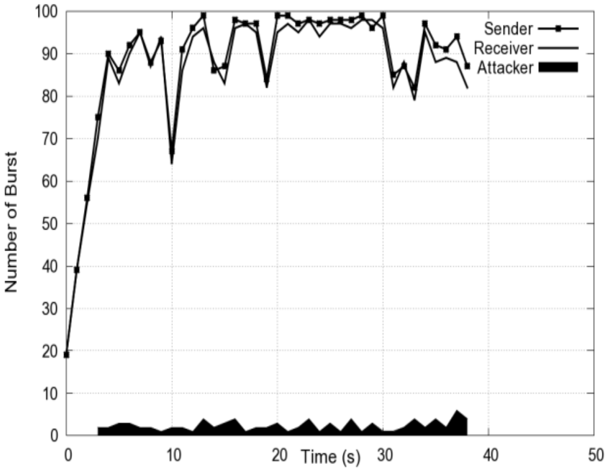


Figure 10. Effect of burst hijacking attack when the number of compromised node is 1.

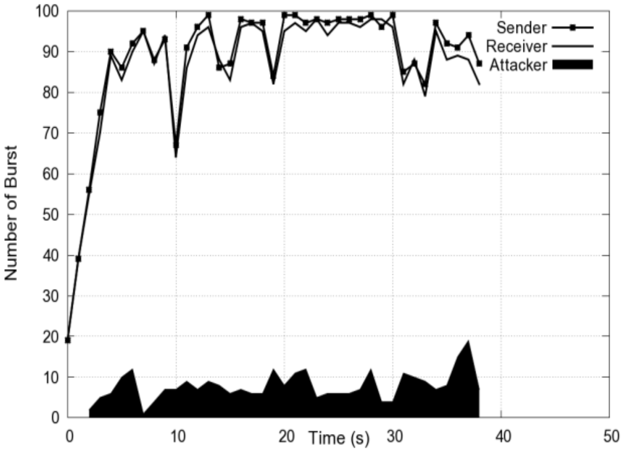


Figure 11. Effect of burst hijacking attack when the number of compromised node is 3.

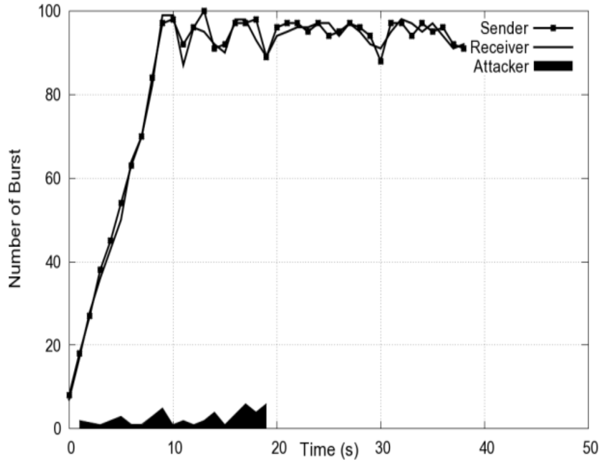


Figure 12. After implementing the solution to burst hijacking attack (number of compromised node = 1)

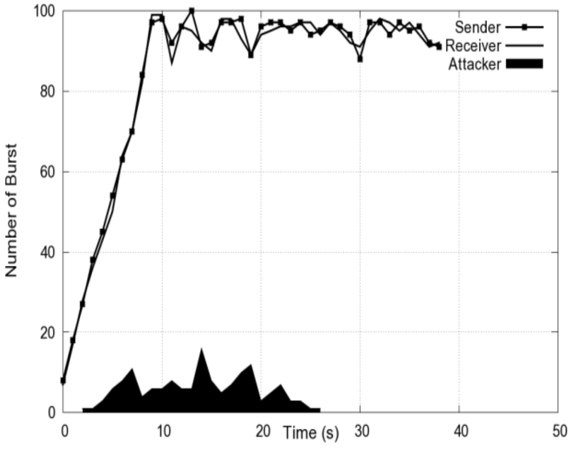


Figure 13. After implementing the solution to burst hijacking attack (number of compromised node = 3)

If the number of compromised nodes is 1 and 3 then the amount of hijacked bursts is also increased as shown in Fig. 10 and Fig. 11. And the Fig 12 and Fig. 13 shows the simulation results after implementing the solution for burst hijacking attack. Even though some bursts are hijacked initially, it is detected by the trusted nodes based on statistical information and an alternate trusted path is used for further communication. Thereby the burst hijacking attack is removed.

VII. Conclusion And Future Work

TCP/OBS networks are the future networks and optical burst switching will turn as the most broadly used technology in the mere future due to its speed and as it provides an end to end optical path among the communicating parties. Since optical burst switching has typical features, it is quite natural to suffer from the security attacks. In this paper, identified the new-fangled type of attack and named as Burst Hijacking Attack. From the statistical approach, its countermeasures are discussed from the normal scenario, attack scenario and attack removal scenario separately using ns2 simulator with the modified nOBS patch.

In the future when the optical burst switching is employed in everywhere then some more security attacks will arise. Future research in this area will help us to identify and remove other possible attack in TCP/OBS networks and make optical burst switching technique a superior one for optical internet.

Acknowledgment

The authors would like to thank the anonymous reviewers and the Editor – in – Chief for their valuable comments that have helped us to improve the manuscript.

References

[1] B. Mukherjee, “WDM Optical Communication Networks: Progress and Challenges,” IEEE Journal on Selected Areas in Communications, pp.1810-1823, October 2000.

[2] C. Qiao and M. Yoo, “Optical Burst Switching (OBS) - a New Paradigm for an Optical Internet”, Journal of High Speed Networks, pp.69-84, January 1999.

[3] S. Verma, H. Chaskar, and R. Ravikanth, “Optical Burst Switching: A Viable Solution for Terabit IP Backbone,” IEEE Network, pp. 48-53, November/December 2000.

[4] S. Yoo, S. J. B. Yoo, and B.Mukherjee, “All-Optical Packet Switching for Metropolitan Area Networks: Opportunities and Challenges,” IEEE Communications Magazine, vol. 39, pp. 142-148, March 2001.

[5] X. Cao, J. Li, Y. Chen, and C. Qiao, “Assembling TCP/IP Packets in Optical Burst Switched Networks., Proceeding of IEEE Globecom, December 2002.

[6] Guray Gurel and Ezhan Karasan, “Effect of Number of Burst Assemblies on TCP Performance in Optical Burst Switching Networks,” Proceedings of the IEEE BROADNETS, October 2006.

[7] M. Yoo and C. Qiao, “A Novel Switching Paradigm for Buffer-Less WDM Networks,” Optical Fiber Communication Conference (OFC), pp. 177-179, February 1999.

[8] J. P. Jue and V. M Vokkarane, “Optical Burst Switching,” Springer

[9] Science, 2005.

[10] M. Yoo and C. Qiao, “Choices, Features and Issues in Optical Burst Switching (OBS),” Optical Networking Magazine, vol. 1, pp. 36-44, April 1999.

[11] C. Siva Ram Murthy and Mohan Gurusamy, “WDM Optical Networks: Concepts, Design and Algorithms,” Prentice Hall PTR, November 2001.

[12] Pushpendra Kumar Chandra, Ashok Kumar Turuk, and Bibhudatta Sahoo, “Survey on Optical Burst Switching in WDM Networks,” Proceding of IEEE communications magazine, December 2009.

[13] Malathi Veeraraghavan and Tao Li, “Signaling Transport Options in GMPLS Networks: In-band or Out-of-band,” International Conference on Computer Communications and Networks, pp. 503-509, August, 2007.

[14] Yuhua Chen and Pramode K. Verma, “Secure Optical Burst Switching: Framework and Research Directions,” IEEE Communication Magazine, pp. 40-45, August 2008.

[15] Yuhua Chen, Pramode K. Verma, and Subhash Kak, “Embedded Security Framework for Integrated Classical and Quantum Cryptography Services in Optical Burst Switching Networks,” Security and Communication Networks, vol. 2, no. 6, pp. 546-554, November-December 2009.

[16] N. Sreenath, G. Mohan and C. Siva Ram Murthy, “Virtual Source Based Multicast Routing in WDM Optical Networks,” IEEE International Conference on Networks (ICON 2000), pp. 385-389, Singapore, September 2000.

[17] [16] Guray Gurel, Onur Alparslan and Ezhan Karasan, “nOBS: an ns2 based simulation tool for performance evaluation of TCP traffic in OBS networks,” Annals of Telecommunications, vol. 62, no. 5-6, pp. 618-632, May-June 2007.

[18] J. Turner, “Terabit Burst Switching,” Journal of High Speed Networks, vol.8, pp. 3-16, January 1999.

[19] Turuk, A. K., Kumar, R., “A Novel Scheme to Reduce Burst-Loss and Provide QoS in Optical Burst switching Network, ” In proceeding of HiPC-2004, pp. 19-22, 2004.

[20] [19] Dolzer. K., Gauger C., Spath J., and Bodamer S.,” Evaluation of reservation mechanisms for optical burst switching ”, AEU International Journal of Electronics and Communications, vol. 55, no. 1, pp. 18-26 April 2001.

[21] Siva Subramanian, P., Muthuraj K.,” Threats in Optical Burst Switched Network. Int. J.Comp. Tech. Appl. “, vol. 2, no. 3, pp. 510-514, July 2011.

[22] N. Sreenath, K. Muthuraj, and P. Sivasubramanian , “ Secure Optical Internet:Attack Detection and Prevention Mechanism,’’ International Conference on Computing, Electronics and Electrical Technologies, 2012.

AUTHORS PROFILE

K. Muthuraj is a Research Scholar and pursuing a Doctoral Degree in Computer science and Engineering at the Department of Computer science and Engineering at Pondicherry Engineering College, Pillaichavady, Puducherry – 605014, India. He received his B.E in Computer science and Engineering (2000) from Madurai Kamaraj University, Madurai, Tamilnadu, India. He received his M.E in Computer science and Engineering (2008) from Anna University, Chennai, Tamilnadu. His research areas are high speed networks and Optical Internet..

Dr. N. Sreenath is a professor and Head of the Department of Computer science and Engineering at Pondicherry Engineering College, Pillaichavady, Puducherry – 605014, India. He received his B.Tech in Electronics and Communication Engineering (1987) from JNTU College of Engineering, Ananthapur – 515002, Andra Pradesh, India. He received his M.Tech in Computer science and Engineering (1990) from University of Hyderabad, India. He received his Ph.D in Computer science and Engineering (2003) from IIT Madras. His research areas are high speed networks and Optical networks.

Qos Routing Scheme and Route Repair in WSN

[1,2]M. Belghachi
[1]Faculty of Science and Technology
University of Bechar,
Bechar, Algeria

[2]M. Feham
[2]STIC Laboratory
Faculty of Technology, University of Tlemcen,
Tlemcen, Algeria

Abstract— **During the last decade, a new type of wireless network has evoked great interest among the scientific community; it is the wireless sensor networks (WSN). The WSN are used in various social activities, such as industrial processes, military surveillance applications, observation and monitoring of habitat, etc... This diversity of applications brings these networks to support different types of traffic and to provide services that must be both generic and adaptive for applications, the properties of the quality of service (QoS) are different from one application to another. However, the need to minimize energy consumption has been the most important field of WSNs research. Few studies in the field are concerned with mechanisms for efficiently delivering QoS at the application level from network level metrics and connection such as delay or bandwidth, while minimizing the energy consumption of sensor nodes that are part of network. The idea is to ensure QoS through a routing process, which can detect paths that meet the QoS requirements based on ant colony optimization (ACOs), coupled with detected routes reservation process. However, it is necessary to integrate to this diagram the maintenance of route disrupted during communication. We propose a method that aims to improve the probability of success of a local route repair. This method based on the density of nodes in the vicinity of a route, as well as on the availability of this vicinity. Taking into account these parameters in the route selection phase (end of the routing process) allows selecting among multiple routes, the one which is potentially the most easily repairable. In addition, we propose a method for early detection of the failure of a local route repair. This method can directly trigger a process of global re-routing that better fits to restore communication between the source and destination.**

Keywords- WSNs; Quality of service; ACO; availability; global re-routing; local re-routing.

I. Introduction

The wireless sensor networks (WSNs) are considered a special type of ad hoc networks. The nodes of such networks consist of a large number of sensor nodes, which can collect and transmit environmental data in an autonomous manner [1, 2]. The position of these nodes is not necessarily predetermined. They are scattered randomly across a geographical area called acquisition field, which defines the field of interest for the phenomenon captured. The data collected are forwarded to a node considered as a collection point, better known as the sink node. WSNs must be dynamically adaptive to continuous changes of the state of the radio channel and must at the same time attempt to satisfy the QoS requirements of the supported applications.

II. Qos Routing In Wsns

Routing is often considered to seek the shortest route in terms of distance between source and destination to transfer data. In the case of routing with quality of service, the goal is not just to find the best path according to some criterion, but to find the best eligible path as well [3]. For this reason, a number of constraints are imposed on the routes in order to determine their eligibility. For example, you may want to search for a route with a certain amount of bandwidth for video traffic or route ensuring that packets are received by the destination within a certain period of time after their issuance by the source. Any route satisfying a quantitative criterion can be described as route ensuring a certain quality of service. WSNs mainly work with the client-server mode. All nodes then attempt to send the captured information to a processing center. This technique is highly energy costly because the information must traverse the entire network to reach the processing center.

III. Acos And Wsn Routing

A new routing approach in WSNs has emerged: this approach is based on algorithms inspired from ant colonies [4, 5]. These algorithms are based on the ability of simple ants to solve complex problems through cooperation. All methods using this paradigm are now called ACOs. Indeed, collective intelligence in social insects results in the emergence of collective behavior due to macroscopic smart simple interactions at the microscopic level. Operation of ant colonies is the best example [6]. The behavior of ants is a collaborative and collective one. Each ant has priority for the welfare of the community. Each individual of the colony is a priori independent and is not supervised (completely distributed system). The colony is self-controlled through relatively simple mechanisms to study. By projecting the behavior of these insects on the characteristics of WSNs, we note that the behavior of ants is well suited to this type of networks, especially when calculating routes.

IV. Description Of The Proposed Approach

The idea is to design a decentralized algorithm based on operation of ants [7, 8], which use their natural ability to find the shortest path between a source and destination while moving through the network. Our approach uses the same mechanisms for selecting local vicinity of Ant System [9, 10].

So, the formula to choose next neighbor with joint attraction pheromone trails. The proposed approach consists of two phases, namely route discovery and route maintenance. When a source node needs to send data to the base station with

QoS requirements (bandwidth), it begins with the route discovery phase. Once the route is found, data transfer can begin. During data transmission, it is also necessary to maintain the path to the destination. One of the weaknesses of these networks lies in the fact that the routes used between source and destination is likely to break surreptitiously during communication. This failure is due to the fact that the nodes forming the route may distrust (deplete their energy). In case the link to the next node is broken, the node initiates phase of route repair. This phase is based on global or local re-routing. The objective of the modification is to ensure the selection of the route more easily repairable among those released during the discovery phase of routes. To achieve this goal, we take into account the character of the nodes vicinity in the network, and in particular the density of nodes, as well as their availability. Repair of route in case of a node failure occurs by the implementation of procedure for local re-routing, and thus avoiding global re-routing that consumes bandwidth and execution time, thereby increasing communication delays.

A. *Interpretation of the availability*

We define the availability of a node by the number of neighbors (ie, the number of nodes in the area of its radio range) whose bandwidth is greater than that required by the connection. The availability parameter is specific to a node and a connection (a given bandwidth). Thus, upon receiving a route request, each node assesses its availability by counting the number of neighbors whose bandwidth is greater than that required by the connection [11]. If the availability is equal to one, no neighbors other than the transmitter of the request meet the QoS requirements of the connection. Continue broadcasting the message of connection is useless because no node in the vicinity is able to extend the route to the destination. This leads to early detection of imminent failure of the route discovery through this node, while reducing network congestion. If the availability is equal to two, one neighbor other than the transmitter of the request meets the QoS of the connection. We can continue to broadcast request message. However, no node will take over from the current node fails in case the route would be used for communication. This availability is too low to ensure a local re-routing of the current node. This is called throttling.

B. *Control packets*

to implement our proposed approach, four control packets are used [12, 13].

1) Hello_Ant packet: The Hello_Ant packet is distributed periodically to all neighbors of the current node, containing the delay of its departure. When neighbors receive this packet, they react by responding by an acquired reception (ACK_Ant). Based on the delay of departure, delay of arrival and the Hello_Ant packet size, the current node calculates the available bandwidth on the links. An entry in the neighbor table is created containing the value of the available bandwidth and the residual energy of all its neighbors. For the update of the value of the bandwidth and the residual energy is produced by subsequent Hello_Ant packets to indicate the current status of links.

2) Route request packet: A Route_Request_Ant packet is broadcast on receiving a route request to a destination with a demand for quality of service expressed in terms of bandwidth. At each node, the hop count is incremented and the node ID is added to the stack of visited nodes.In addition to the exploration of the shortest path between the source and the destination, Route_Request_Ant collect the end-to-end delay, the minimal available bandwidth, the average availability and the minimal residual energy of the path through which it is propagated.

3) *Route reply packet*: upon receiving Route_Request_Ant, the destination creates the response message Route_Reply_Ant. Route_Request_Ant packet collects the transmission delay of each link, the processing time at each node, the bandwidth available on each link, residual energy and hop count. This Route_Reply_Ant will be sent (unicast) to the original source along the route established by route_Request_Ant in reverse.

4) Route error packet: This Route_error packet is sent to the source of the communication to indicate that the route to a destination is broken and that the attempt of local re- routing was unsuccessful. Local repair is attempted only on condition that the availability of node is strictly greater than two. When the source receives this packet, it invalidates the status corresponding to the destination, stops sending packets to the destination and places them in a queue. Meanwhile, the source initiates a procedure of global re-routing.

C. *Mathematical Model*

The objective function of the proposed work is to find a path from source to destination through a neighbor with a maximum transition probability. The probability of transition from source i to destination d through neighbor j of i is calculated as follows [14, 15]:

$$P_{ijd} = \frac{\left[Dispo_{ijd}\right]^{\alpha}\left[D_{ijd}\right]^{\beta}\left[\eta_{ijd}\right]^{\chi}\left[BP_{ijd}\right]^{\delta}}{\sum_{l\in N_i}\left[Dispo_{il}\right]^{\alpha}\left[D_{ild}\right]^{\beta}\left[\eta_{ild}\right]^{\chi}\left[BP_{ild}\right]^{\delta}} \qquad (1)$$

Where α, β, γ and δ (> = 0) are parameters that control the relative importance between availability, delay, residual energy / hop count and available bandwidth. Ni is the set of neighbors of i and l is neighbors of i through which a route is available to the destination. For the calculation of relatives' metrics, the delay and the number of hops are additives metric; bandwidth and the residual energy are considered as non-additive concave metric. Additive metrics must be reduced to a minimum for the shortest paths; the non-additive concave metric is used to maximize bandwidth and residual energy [16, 17].

1) *Availability:*

The problem is to determine the extent to which a node is part of a route between a source and destination and immediately adjacent to another node (i.e. its radio range) and therefore can be replaced by the latter in case of failure. It is particularly important that the replacement node have enough bandwidth communication channels for this new connection.

$$\begin{aligned} &Dispo_{ijd} = Moy\{Availibility(l)\} \\ &\square\forall\, l \in route_j(i,d) \end{aligned} \qquad (2)$$

Dispoijd: the average number of neighbors whose bandwidth is greater than that required by the connection along the path from i to d through j.

2) Delay :

The delay between the source and the destination is calculated by:

$$D_{ijd} = \sum_{l \in route_j(i,d)} delay(l) \qquad (3)$$

Where the delay (l) is the end to end delay from source i to destination d through the neighbor j by route request message at the time of route exploration.

3) hops count / Minimum Residual Energy :

This relative metric equal the inverse of number of hops multiplied by the minimum Residual battery Energy of all intermediate nodes between source and destination is given by:

$$\begin{aligned} &MBR_{ijd} = \min\{\mathrm{Re}\,sidual\_Energy(l)\} \\ &\forall\, l \in route_j(i,d) \end{aligned} \qquad (4)$$

$$\eta_{ijd} = \frac{MBR_{ijd}}{NbSaut(route_j(i,d))} \qquad (5)$$

Where

- MBRijd: minimum residual energy along the path i to d through j
- NbSaut (routej (i, d)) is the number of hops along the path i to d through j.

4) The bandwidth:

Available bandwidth on the path from i to d is calculated as the minimum of available bandwidth (Bpijd) of all links along that path.

$$\begin{aligned} &Bp_{ijd} = \min\{Available\_bandwith(l)\} \\ &\forall\, l \in route_j(i,d) \end{aligned} \qquad (6)$$

Hello_Ant messages are often transmitted to keep the connectivity of the vicinity, and they can better reflect the current available bandwidth on the links rather than the route search messages.

D. Routes discovery phase

The source initiates the routing process. It sends to all its neighbors a connection request (Route_Request_Ant) to the destination with QoS requirements in terms of bandwidth. Nodes that receive the message for the first time and meet the QoS requirements broadcast a demand to their vicinity after collecting:

a) The transmission delay of each link

b) The available bandwidth on each link

c) The number of hop

d) The residual energy

This connection message is thus spread in the network until it reaches the destination. When Route_Request_Ant reaches the destination, it will be converted into Route_Reply_Ant and transmitted to the origin source. Route_Reply_Ant takes the same path marked by Route_Request_Ant in reverse. For each Route_Reply_Ant, when reaching an intermediate node or source node, the node just find the delay, bandwidth, residual energy and the number of hops. The node can calculate the probability to reach the destination. The path with the highest probability is considered as the best, and the data transmission can begin.

E. Route maintenance phase

Because of the failure of nodes responsible for the transmission of data between source and destination, the risk that the route be disrupted before the end of the communication is very high. In case of link-breaking or node failure during data transmission, there are two scenarios for the re-routing [11, 18]:

1) Global re-routing from the source of the communication. This re-routing is implemented in most routing protocols, although it takes a lot of time and consumes a lot of bandwidth.

2) Local re-routing from the node where the failure occurred. This local routing has the advantage of being fast and consume less bandwidth. Local repair is attempted only on condition that the availability is strictly greater than two. Otherwise, a Route error is sent directly towards the source of the communication to undertake a global repair of the route. We bring this change to avoid loss of time expected from a routing attempt to local route repair anyway doomed to failure (if availability is less than three) and thus accelerate the re-routing phase despite hostile conditions.

F. Experimental design

In our approach, we seek to exploit the heterogeneity of nodes distribution in the network by taking into account the availability parameter. It would be pointless to base our tests on a random configuration improvement that is statistically homogeneous. Nevertheless, we conducted several simulations in such a configuration to conclude that our contribution adds little performance compared to AODV protocol. To perform our tests, we chose a concentrate on a heterogeneous configuration of data, and then to change the departure of nodes. In each simulation, we consider the following cases:

a) Case 1: AODV protocol.

b) Case 2: Our approach taking into account the availability parameter and re-routing based on local density.

Different results are presented in the following section and correspond to averages over a series of simulations. In all scenarios the nodes are static, and randomly deployed. Parameter values of relative weight α, β, γ, δ and δ are defined in Table I

Table I. Parameter settings

Parameters	Values
Availability weight, α	1.0
Delay weight, β	1.0
Residual Energy/ hop count weight, γ	1.0
Available bandwidth weight, δ	1.0
Interval Hello_Ant	1 sec
Hello_Ant Retry times	3

G. Tests and analysis of results

In the configuration shown in Figure 1, node 2 tries to establish communication with the Sink, thus triggering the route research phase to the destination by broadcasting a Route_Request_Ant packet with QoS requirements in terms of bandwidth. Several types of routes are possible. Among these, the route containing nodes 3, 4, 5 and 6 is the shortest (Case 1). To test our hypotheses, we plan the "disappearance" of a node at a specific time. Then we focus our interest in the evolution of the system and in particular the way to repair the route. We decide to envisage the disappearance of nodes 6, 13 and 18. This choice leans on the fact that the study of the disappearances of the other nodes is impossible (disappearance of nodes 2 and/or Sink which would hinder definitively the communication).

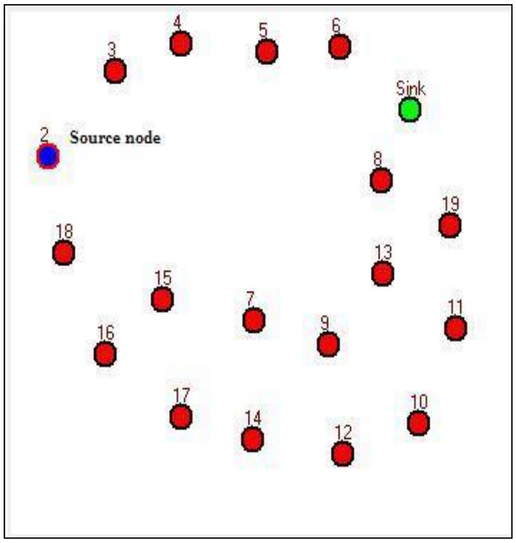


Figure 1. Spatial distribution of nodes

So, the node 7 receives a Route_Request_Ant packet for the destination passed through the nodes 18 and 15. The node 7 adds in its database an entry for the destination by indicating that the next node for the destination is 15. Then, the node 7 updates the values of average availability, End to End delay, Residual energy and hop count. The Route_Request_Ant packet can be then rebroadcasted.

The first packet which arrives at the destination is the packet passed through the nodes 3, 4, 5 and 6. A time-out whose value depends on the delay taken by the packet to propagate until the destination is immediately launched. In this interval, the destination stores each Route_Request_Ant packet that arrives. So, another Route_Request_Ant has passed through the nodes 18, 15, 7, 9, 13 and 8 (Case 2) is treated by the destination. The length of the detected route is equal to 7.

1) Disappearance of node 6:

Table II summarizes the results obtained during the simulations. In case1, we notice that time and the data are strongly influenced by the disappearance of node 6. The regular delay that averages 9.5ms before the disappearance of node 6 rises to 14 ms and the route used extends in terms of number of nodes. These observations can be explained by the fact that this node is part of the route used to transport data to the destination when it disappears. Node 5 notices that the node 6 has disappeared and immediately returns a route errors message to the source node. This triggers an operation of global re-routing. As shown in Table II, the re-routing phase lasts longer than the initial phase (18ms against 9.2ms).

In case 2, the packets follow a path in which node 6 is not part of the route used. So we do not see any change. Delays remain unchanged before and after the departure of node 6. It establishes at 14ms. By contrast, we note that the initial route research delay is much longer than in the previous case. We explain this observation by the fact that we expect more Route_Request_Ant at the destination before returning Route_Reply_Ant to destination on the selected route.

Table II. Summary of results when node 6 disappears

	Case 1	Case 2
Initial routing delay (millisecond)	9.2	19.729
local repair delay (millisecond)	0	0
total re-routing delay (millisecond)	18.1667	0
Average delay before re-routing (millisecond	9.5	14
Average delay after re-routing (millisecond)	16	14
Average length of routes (node)	6.563	7

2) Disappearance of node 13:

The data collected for case 1 indicates that no re-routing is undertaken following the departure of node 13. We could predict it because the node 13 is not part of the route used in case 1. In case 2 however, we notice that the results are strongly affected. Thus, the delay increases from 14ms before the disappearance of node 13 to 16ms. We also observe in Table III that local re-routing was undertaken successfully. We note that the local re-routing lasts 7.82ms. This re-routing delay is compared to the 18ms for global re-routing in case 1 obtained for the disappearance of node 6.

Table III. Summary of results when node 13 disappears

	Case 1	Case 2
Initial routing delay (millisecond)	8.55	22.013
local repair delay (millisecond)	0	7.82
total re-routing delay (millisecond)	0	7.82
Average delay before re-routing (millisecond)	9.5	14
Average delay after re-routing (millisecond)	9.5	16
Average length of routes (node)	5	8.261

3) Disappearance of node 18:

As for the disappearance of node 13, only case 2 is affected by the disappearance of node 18. Moreover, we observe in Table IV. That the delay decreases 14ms before the disappearance of node to 9.5ms after its disappearance. These results for case 2 are to be compared with those obtained for case 1 in the disappearance of node 3. We note that the length of the search path following the disappearance of node 18 is 39ms. This delay is much less than the re-routing duration in case 1 measured for the departure of node 3.

Table IV. Summary of results when node 18 disappears

	Case 1	Case 2
Initial routing delay (millisecond)	8.55	19.756
local repair delay (millisecond)	0	39.012
total re-routing delay (millisecond)	0	30.12
Average delay before re-routing (millisecond)	9.5	14
Average delay after re-routing (millisecond)	9.5	9.5
Average length of routes (node)	5	5.937

4) Assessment of node disappearance:

We seek to extend our results to the disappearance of nodes 3, 6, 13 and 18 to all network nodes. Thus, we consider the disappearance of all nodes belonging to an initial route. We do not envisage the disappearance of node 2 and/or Sink (their disappearance would be unrecoverable for communication) as for nodes 10, 11, 12, 14, 17, 16 or 19. The disappearance of the latter in fact has a limited impact in the system. We assume that the disappearance of each node in the system is equi-probable. We classify nodes into several groups according to the case to which we refer. We obtain the following distributions:

- Case 1: The disappearance of node 13 is representative of the disappearance of nodes as follows: 7, 8, 9, 13, 15 and 18. The disappearance of node 6 is representative of the disappearance of nodes 4, 5 and 6.
- Case 2: The disappearance of node 6 is representative of the disappearance of nodes: 3, 4, 5 and 6. The disappearance of node 14 is representative of the disappearance of nodes 7, 8, 9, 13 and 15.

Using these distributions, we weight the different scenarios for the disappearance synthetic results presented in Table V

Table V. Results of Case1 and Case2 comparison

	Cas 1	**Cas 2**
Initial routing delay (millisecond)	8.78	20.5
total re-routing delay (millisecond)	39	5.8
Average delay before re-routing (millisecond	9.5	14
Average delay after re-routing (millisecond)	11.9	14.6
Average length of routes (node)	5.63	7.52

We note that the initial routing delay is always longer for case 2. This is the main drawback inherent in our improvement. However, waiting longer at the route establishment phase is necessary to choose a better repairable route. We also note that the average re-routing caused by the departure of a node in case 2 is significantly better than case 1. Thus, the waiting time is on average 6.7 times shorter than in case 1. This observation attests to the significant improvement of the QoS management offered by the protocol improved compared to the original version. A weakness of our improvement is observed in the average length of the route. We find that the route is on average equal to 7.52 in Case 2 against about 5.5 for the other case. The route thus involves more nodes and, therefore, the probability that a failure occur increases on the route retained by case 2. Another weakness of our improvement is that the end to end delay of data packets is more important in our version compared to the original version. In case 2, the end to end delay is established before the disappearance 9.5ms against 14ms for both cases. By contrast, we find that, this delay increases shortly after the disappearance and then settled at 14.6ms against 11.9ms for case 1. The end to end delay obtained in case 2 is more stable than that observed in the other case. We conclude that this level, once again our improvement provides a gain in terms of the QoS management in case of node disappearance.

V. Conclusion

The work presented in this paper aims to improve the QoS management by taking into account the availability defined as the number of available nodes in the radio range of a node. Our approach was based on a thorough analysis of the tools taking into account the QoS in WSN. We then devised the concept of availability and describe how a node can use this information to improve QoS. We have identified two main areas:

- Establishment of a mechanism for route choice.
- Prediction of early failure of a local re-routing.

The mechanism of route choice is to select among several competitors whose maintenance is the easiest to achieve. Simulation results confirm our theoretical reasoning: the delay of re-routing is improved. In addition, our mechanism also improves the rate of packet loss. QoS is half-open thus further strengthened. We have highlighted a situation in which a local re-routing attempt fails: if the availability is too low around the node that initiates local re-routing, it is doomed to failure. It is therefore preferable to return directly to the source a Route error packet after detecting a link failure. The main limitation of our route selection process is the loss of time in the initial routing. This loss of time is related to the expectation of receiving other Route_Request_Ant from the source before selecting the best route to facilitate route repair. Another limitation stems from the fact that the routes adopted after our improvement are longer. The end to end delay is, therefore more important. Another line of work would be to conduct a thorough performance study for distributions of nodes inspired from plausible situations in real environments.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, “Wireless Sensor Networks: A Survey,” Computer Networks, Vol. 38, No. 4, March 2002, pp. 393-422.

[2] H. Karl and A. Willing, “A Short Survey of Wireless Sensor Networks,” Technical Report TKN-03-018, Telecommunication Networks Group, Technical University, Berlin, October 2003.

[3] D. Chen and P. K. Varshney, “QoS Support in Wireless Sensor Network: A Survey,” Proceedings of the 2004 International Conference on Wireless Networks (ICWN2004), Las Vegas, Nevada, USA, June 2004.

[4] M. Dorigo, “Optimization, learning and natural algorithms (initalian),” Ph.D. dissertation, 1992.

[5] R. Montemanni and L. Gambardella, “Swarm approach for a connectivity problem in wireless networks,” in Proceedings 2005 IEEE Swarm Intelligence Symposium, pp. 265—272, 2005.

[6] M. Dorigo and T. Stuzle, Ant Colony Optimization. MIT Press, 2004.

[7] R. Montemanni and L. Gambardella, “Swarm approach for a connectivity problem in wireless networks,” in Proceedings 2005 IEEE Swarm Intelligence Symposium, pp. 265—272, 2005.

[8] W. Agassounon, “Distributed information retrieval and dissemination in swarm-based networks of mobile, autonomous agents,” in Proceedings of IEEE Swarm Intelligence Symposium, pp. 152—159, 2003.

[9] O. Cordon, F. Herrera and T. Stuzle, “A review on the ant colony optimization metaheuristic: Basis, models and new trends,” Mathware and software computing, no. 9, pp. 1—35, 2002.

[10] P. Ji, et al. “DAST: A QoS-Aware Routing Protocol for Wireless Sensor Networks,” Proceeding of International Conferences on Embedded Software and Systems Symposia, Sichuan, 29-31 July 2008, pp. 259-264.

[11] M.Belghachi, M.Feham,“QoS Based on Ant Colony Routing for Wireless Sensor Networks,” International Journal of Computer Science and Telecommunications, January 2012.

[12] M.Belghachi, M.Feham, “Routing diagram for the transport of video traffic in a WSN,” Journal of Theoretical and Applied Information Technology, July 2010.

[13] P Deepalakshmi, S Radhakrishnan, Ant colony based QoS routing algorithm for mobile ad hoc networks. Int J Recent Trends Eng. 1(1), 459–462 (2009).

[14] M Gunes, U Sorges, I Bouazzi, ARA–the ant-colony based routing algorithm for MANETs, in Proceedings of the International Conference on Parallel Processing Workshops (ICPPW'02), Vancouver, BC (2002).

[15] O Hussein, T Saadawi, Ant routing algorithm for mobile ad hoc networks (ARAMA), in Proceedings of the International Performance Computing and Communications Conference (IPCCC 2004), Phoenix, Arizona (2004).

[16] G Di Caro, F Ducatelle, LM Gambardella, AntHocNet: an adaptive nature inspired algorithm for routing in mobile ad hoc networks. Eur Trans Telecommun. 16(2), 443–455 (2005).

[17] Z Liu, MZ Kwiatkowska, C Constantinou, A biologically inspired QoS routing algorithm for mobile ad hoc networks. Int J Wirel Mobile Comput. 4(2), 64–75 (2010).

[18] B. Macabéo " Routage et reparation de routes dans les réseaux mobiles Ad hoc", Maitrise en genie électrique Avril 2003.

A Novel Feistel Cipher Involving a Bunch of Keys supplemented with Modular Arithmetic Addition

Dr. V.U.K Sastry
Dean R&D, Department of Computer Science and Engineering, Sreenidhi Institute of Science & Tech.
Hyderabad, India

Mr. K. Anup Kumar
Associate Professor, Department of Computer Science and Engineering, Sreenidhi Institute of Science & Tech.
Hyderabad, India

Abstract— **In the present investigation, we developed a novel Feistel cipher by dividing the plaintext into a pair of matrices. In the process of encryption, we have used a bunch of keys and modular arithmetic addition. The avalanche effect shows that the cipher is a strong one. The cryptanalysis carried out on this cipher indicates that this cipher cannot be broken by any cryptanalytic attack and it can be used for secured transmission of information.**

Keywords- encryption; decryption; cryptanalysis; avalanche effect; modular arithmetic addition.

I. INTRODUCTION

In the development of block ciphers in cryptography, the study of Feistel cipher and its modifications is a fascinating area of research. In a recent investigation [1], we have developed a novel block cipher by using a bunch of keys, represented in the form of a matrix, wherein each key is having a modular arithmetic inverse. In this analysis, we have seen that the multiplication of different keys with different elements of the plaintext, supplemented with the iteration process, has resulted in a strong block cipher, this fact is seen very clearly by the avalanche effect and the cryptanalysis carried out in this investigation.

In this paper, we have modified the block cipher developed in [1] by replacing the XOR operation with modular arithmetic addition. Here our interest is to study how the modular arithmetic addition influences the iteration process and the permutation process involving in the analysis.

In what follows, we present the plan of the paper. In section 2, we deal with the development of the cipher and introduce the flow charts and the algorithms required in this analysis. We have illustrated the cipher in section 3, and depicted the avalanche effect. Then in section 4, we carry out the cryptanalysis which establishes the strength of the cipher. Finally, we have computed the entire plaintext by using the cipher and have drawn conclusions obtained in this analysis.

Development Of The Cipher

Consider a plaintext containing 2m2 characters. Let us represent this plaintext in the form of a matrix P by using EBCIDIC code. We divide this matrix into two square matrices P0 and Q0, where each one is matrix of size m.

The equations governing this block cipher can be written in the form

$$[\ P_{jk}^{\ i}\ ] = [\ e_{jk}\ Q_{jk}^{\ i-1}\ ] \bmod 256, \qquad (2.1)$$

and

$$[\ Q_{jk}^{\ i}\ ] = ([e_{jk}\ P_{jk}^{\ i-1}] \bmod 256 + [Q_{jk}^{\ i-1}]) \bmod 256\ , \qquad (2.2)$$

where j= 1 to m , k = 1 to m and i =1 to n, in which n is the number of rounds.

the equations describing the decryption are obtained in the form

$$[\ Q_{jk}^{\ i-1}\ ]= [\ d_{jk}\ P_{jk}^{\ i}\ ] \bmod 256, \qquad (2.3)$$

and

$$[\ P_{jk}^{\ i-1}\ ]= [d_{jk}(\ [\ Q_{jk}^{\ i}\ ]\ -\ [\ Q_{jk}^{\ i-1}\ ]\ )\ ]\ \bmod 256 \qquad (2.4)$$

where j= 1 to m , k = 1 to m and i = n to 1,

Here ejk , j = 1 to m and k = 1 to m, are the keys in the encryption process, and djk j = 1 to m and k = 1 to m, are the corresponding keys in the decryption process. The keys ejk and djk are related by the relation

$$(\ e_{jk}\ d_{jk}\ ) \bmod 256 = 1, \qquad (\ 2.5)$$

that is, d_{jk} is the multiplicative inverse of the given e_{jk} . Here it is to be noted that both e_{jk} and d_{jk} are odd numbers which are lying in [1-255].

For convenience, we may write

$$E = [\ e_{jk}\ ]\ , \qquad j = 1 \text{ to } m \quad \text{and} \quad k = 1 \text{ to } m.$$

and

$$D = [\ d_{jk}\ ]\ , \qquad j = 1 \text{ to } m \quad \text{and} \quad k = 1 \text{ to } m.$$

where E and D are called as key bunch matrices.

The flow charts describing the encryption and the decryption processes are given by

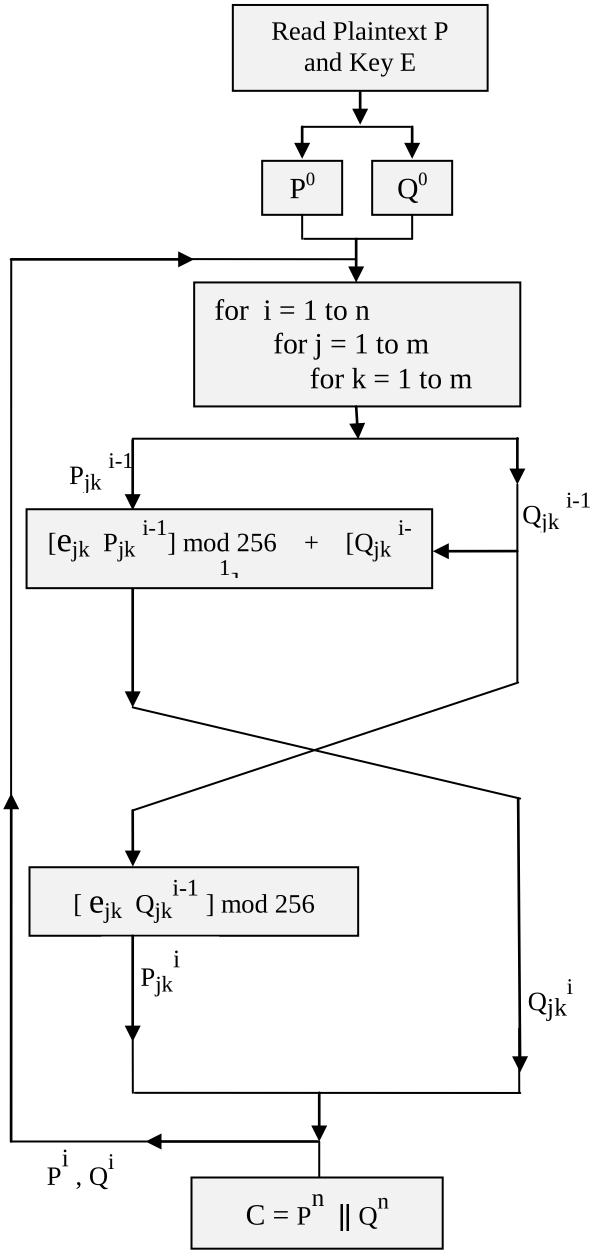


Figure 1. The Process of Encryption

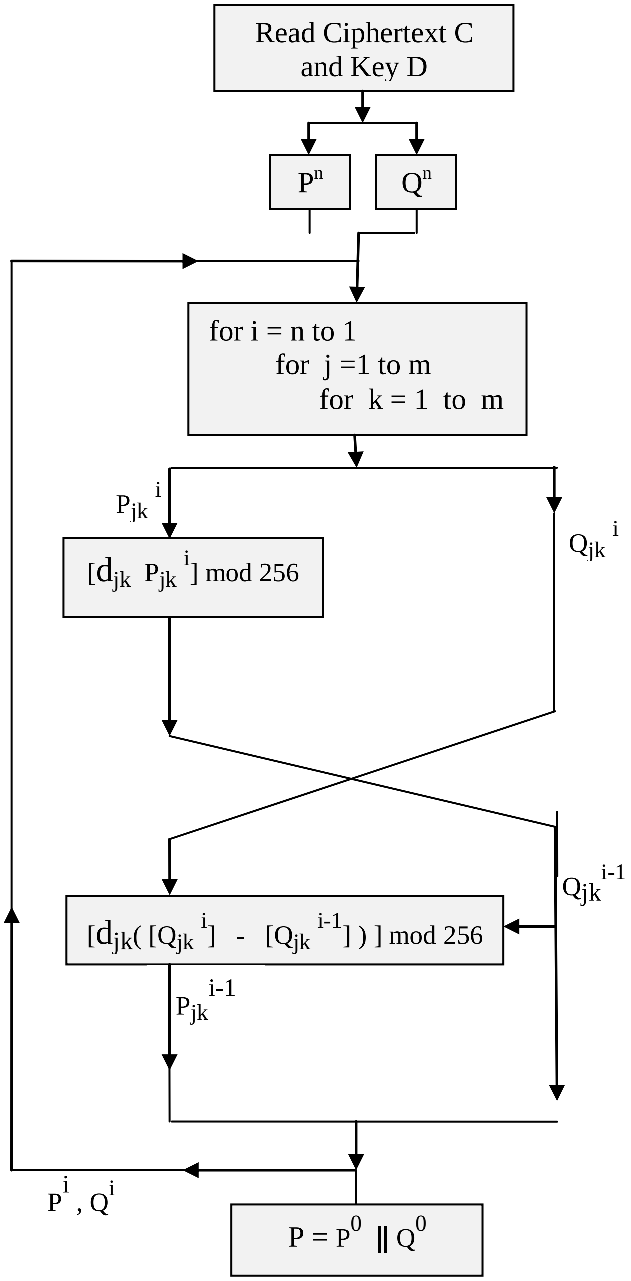


Figure 2. The process of Decryption

The corresponding algorithms are written in the form given below.

A. Algorithm for Encryption

```
1.  Read P, E, and n
2.  $P^0$ = Left half of P.
    $Q^0$ = Right half of P.
3. for i = 1 to n
   begin
       for j = 1 to m
       begin
           for k = 1 to m
             begin
               $[P_{jk}^{i}] = [e_{jk} Q_{jk}^{i-1}] \bmod 256$,
               $[Q_{jk}^{i}] = [e_{jk} P_{jk}^{i-1}] \bmod 256 + [Q_{jk}^{i-1}]$ ,
             end
       end
   end
6. C  = $P^n \| Q^n$  $\|$ /*   represents concatenation  */
7. Write(C)
```

B. Algorithm for Decryption

```
1. Read C, D,  and  n.
2. $P^n$ = Left half of C
   $Q^n$ = Right half of C
3.  for i = n to 1
begin
    for j = 1 to m
       begin
           for k = 1 to m
             begin
             $[Q_{jk}^{i-1}] = [d_{jk} P_{jk}^{i}] \bmod 256$,
             $[P_{jk}^{i-1}] = [d_{jk} ([Q_{jk}^{i}] - [Q_{jk}^{i-1}]] \bmod 256$
            end
       end
end
6. P  = $P^0 \| Q^0$   /*$\|$ represents concatenation */
7. Write (P)
```

II. Illustration Of The Cipher

Consider the plaintext given below

Sister! What a pathetic situation! Father, who joined congress longtime back, he cannot accept our view point. That’s how he remains isolated. Eldest brother who have become a communist, having soft corner for poor people, left our house longtime back does not come back to our house! Second brother who joined Telugu Desam party in the time of NTR does not visit us at any time. Our brother in law who is in Bharathiya Janata Party does never come to our house. Mother is very unhappy! (3.1)

Let us focus our attention on the first 32 characters of the above plaintext. This is given by

Plaintext (3.2)

On using the EBCIDIC code, we obtain

$$P = \begin{bmatrix} 083 & 105 & 115 & 116 & 101 & 114 & 033 & 032 \\ 087 & 104 & 097 & 116 & 032 & 097 & 032 & 112 \\ 097 & 116 & 104 & 101 & 116 & 105 & 099 & 032 \\ 115 & 105 & 116 & 117 & 097 & 116 & 105 & 111 \end{bmatrix} \qquad (3.3)$$

This can be written in the form

$$P^0 = \begin{bmatrix} 083 & 105 & 115 & 116 \\ 087 & 104 & 097 & 116 \\ 097 & 116 & 104 & 101 \\ 115 & 105 & 116 & 117 \end{bmatrix} \qquad (3.4)$$

and

$$Q^0 = \begin{bmatrix} 101 & 114 & 033 & 032 \\ 032 & 097 & 032 & 112 \\ 116 & 105 & 099 & 032 \\ 097 & 116 & 105 & 111 \end{bmatrix} \qquad (3.5)$$

Let us now take the key bunch matrix E in the form

$$E = \begin{bmatrix} 125 & 133 & 057 & 063 \\ 005 & 135 & 075 & 015 \\ 027 & 117 & 147 & 047 \\ 059 & 107 & 073 & 119 \end{bmatrix} \qquad (3.6)$$

On using the concept of multiplicative inverse, given by the relation (2.5), we get the key bunch matrix D in the form

$$D = \begin{bmatrix} 213 & 077 & 009 & 191 \\ 205 & 055 & 099 & 239 \\ 019 & 221 & 155 & 207 \\ 243 & 067 & 249 & 071 \end{bmatrix} \qquad (3.7)$$

On using (3.4) – (3.6) and applying the encryption algorithm, we get the ciphertext C in the form

$$C = \begin{bmatrix} 036 & 138 & 014 & 142 & 000 & 238 & 090 & 106 \\ 110 & 090 & 214 & 104 & 144 & 118 & 246 & 206 \\ 016 & 022 & 098 & 018 & 194 & 218 & 070 & 114 \\ 108 & 120 & 038 & 118 & 208 & 224 & 146 & 196 \end{bmatrix} \qquad (3.8)$$

On using the ciphertext C given by (3.8), the key bunch D given by (3.7), and the decryption algorithm given in section 2, we get back the original plaintext.

Now let us consider the avalanche effect which predicts the strength of the cipher.

On changing the fourth row, fourth column element of P0 from 117 to 119, we get a one bit change in the plaintext as the EBCIDIC codes of 117 and 119 are 01110101 and 01110111. On using the modified plaintext and the encryption key bunch matrix E we apply the encryption algorithm, and obtain the corresponding ciphertext in the form

$$C = \begin{bmatrix} 060 & 106 & 182 & 142 & 076 & 198 & 038 & 132 \\ 182 & 196 & 242 & 196 & 000 & 034 & 194 & 240 \\ 140 & 252 & 088 & 140 & 108 & 090 & 146 & 124 \\ 042 & 022 & 094 & 180 & 156 & 250 & 206 & 084 \end{bmatrix} \quad (3.9)$$

On comparing (3.8) and (3.9) in their binary form, we find that these two ciphertext differ by 129 bits out of 256 bits. This shows the strength of the cipher is quite considerable.

Now let us consider the one bit change in the key, On changing second row, third column element of E from 75 to 74, we get a one bit change in the key. On using the modified key, the original plaintext (3.2) and the encryption algorithm, we get the cipher text in the form

$$C = \begin{bmatrix} 242 & 248 & 202 & 122 & 058 & 004 & 036 & 154 \\ 022 & 252 & 002 & 206 & 104 & 098 & 116 & 002 \\ 190 & 108 & 190 & 072 & 250 & 106 & 022 & 200 \\ 044 & 114 & 220 & 222 & 050 & 106 & 030 & 220 \end{bmatrix} \quad (3.10)$$

On comparing (3.8) and (3.10), in their binary form, we find that these two ciphertexts differ by 136 bits out of 256 bits. This also shows that the cipher is expected to be a strong one.

III.Cryptanalysis

In the literature of the cryptography the strength of the cipher is decided by exploring cryptanalytic attacks. The basic cryptanalytic attacks that are available in the literature [2] are

1) *Ciphertext only attack (Brute Force Attack),*
2) *Known plaintext attack,*
3) *Chosen plaintext attack, and*
4) *Chosen ciphertext attack.*

In all the investigations generally we make an attempt to prove that a block cipher sustains the first two cryptanalytic attacks. Further, we make an attempt to intuitively find out how far the later two cases are applicable for breaking a cipher.

As the key E is a square matrix of size m, the size of the key space is

$$2^{(8m^2)} = (2^{10})^{0.8 m^2} \approx (10^3)^{0.8 m^2} = (10)^{2.4m^2}$$

If we assume that the time required for the encryption with each key in the key space as 10-7 seconds, then the time required for the execution with all the keys in the key space is

$$\frac{10^{(2.4m^2)} \times 10^{-7}}{365 \times 24 \times 60 \times 60} \text{ years} = 3.12 \times 10^{(2.4 m^2 -15)} \text{ years}$$

In the present analysis, as m=4, the time required is given by 3.12 x 10 23.4 years. As this is a formidable quantity we can readily say that this cipher cannot be broken by the brute force approach.

Let us know examine the strength of the known plaintext attack. If we confine our attention to one round of the iteration process, that is if n = 1, the equations governing the encryption are given by

$$[P_{jk}^{1}] = [e_{jk} Q_{jk}^{0}] \bmod 256, \quad (4.1)$$

$$[Q_{jk}^{1}] = [e_{jk} \; P_{jk}^{0}] \bmod 256 + [Q_{jk}^{0}], \quad (4.2)$$

where, j = 1 to m, and k = 1 to m.

and

$$C = P^1 \parallel Q^1 \; . \quad (4.3)$$

In the case of this attack, as C, yielding P_{jk}^{1} and Q_{jk}^{1} and as P yielding P_{jk}^{0} and Q_{jk}^{0} are known to the attacker, he can readily determine e_{jk} by using the concept of the multiplicative inverse. Thus let us proceed one step further.

On considering the case corresponding to the second round of the iteration (n = 2), we get the following equations in the encryption process.

$$[P_{jk}^{1}] = [e_{jk} Q_{jk}^{0}] \bmod 256, \quad (4.4)$$

and

$$[Q_{jk}^{1}] = [e_{jk} \; P_{jk}^{0}] \bmod 256 + [Q_{jk}^{0}], \quad (4.5)$$

$$[P_{jk}^{2}] = [e_{jk} Q_{jk}^{1}] \bmod 256, \quad (4.6)$$

and

$$[Q_{jk}^{2}] = [e_{jk} \; P_{jk}^{1}] \bmod 256 + [Q_{jk}^{1}], \quad (4.7)$$

where, j = 1 to m and k = 1 to m.

Further we have,

$$C = P^2 \parallel Q^2 \; . \quad (4.8)$$

Here P_{jk}^{0} and Q_{jk}^{0} are known to us, as C is known. We also know P_{jk}^{0} and Q_{jk}^{0} as this is the known plaintext attack. But here, we cannot know P_{jk}^{1} and Q_{jk}^{1} either from the forward side or from the backward side. Thus e_{jk} cannot be determined by any means, and hence this cipher cannot be broken by the known plaintext attack.

As the equations governing the encryption are complex, it is not possible to intuitively either a plaintext or a ciphertext and attack the cipher. Thus the cipher cannot be broken by the last two cases too. Hence we conclude that this cipher is a very strong one.

IV. Computations And Conclusions

In this investigation we have developed a block cipher by modifying the Feistel cipher. In this analysis the modular arithmetic addition plays a fundamental role. The key bunch encryption matrix E and the key bunch decryption matrix D play a vital role in the development of the cipher. The computations involved in this analysis are carried out by writing programs in C language.

On taking the entire plaintext (3.1) into consideration, we have divided it into 14 number of blocks. In the last block, we have included 26 blanks characters to make it a complete block. On taking the encryption key bunch E and carrying out the encryption of the entire plaintext, by applying encryption algorithm given in section 2, we get the ciphertext C in the form given below

```
 128  100  202   018  120  154  146   058  148  244  200   026  152  198   056  176
 086  066  184   182  192  178  146   236  224  058  082   198  078  218   060  236
 176  156  224   178  070  200  014   090  078  252  230   042  180  108   090  084
 102  060  144   244  240  184  088   190  150  056  110   254  146  222   006  206
 074  182  128   236  074  024  058   104  242  182  024   140  078  012   184  126
 090   088 194   182  170  096  054   122  058  146  014   028  050  204   036  138
 178   076 130   182  130  028  228   184  146  044  238   056  250  176   224  136
 128  188  188    046 074  076  100   182  014  222  050   134  178  214   228  230
 044  254  210    094 076  0 98 216   036  098  236  238   072  254  090   234  108
 172  022  198    146 028  182  054   140  154  134  182   054  034  182   054  240
 102  048  180    110 076  244  178   014  222  248  226   00 2 204   098  106  122
 090  236  108    170 052  200  058   122  098  026  090   218  242  196   004  106
 176  182  172    138  074 140  230   146  214  198  228   102  250  112   086  104
 124  240   000   246  144 220  116   046  126  250  108   222  206  202   250  048
 000  246   116   238  178 244  134   228  058  206  108   190  144  044   152  098
 078  050   114   102  082 190  152   00 2 0 82  024 198   054  042  232   118  054
140  198    038   134  220 190  044   044  096  218   084  176  026  060   028  200
 134  014   152   230  146 196  088   166  064  218   192  014  114  220   200   022
 246 156    252   216  240 196  064    094 222  150   036  038  050  218   006  110
 152 194    216   234  114 114  150   254  232  046   166  176  108  146   176  118
 246  036   254   044  244  054  214  138  098   072  142   090  154  198   076  066
 218  154   144   090  026  248  178  024  218   182   038  250  088  006  110  124
 240  000   102   048  180  188  172  118  054   212   176  104  080  156  242  070
 214  198   228   102  250   092 228  190  250   074   020  102  152   006 110  076
 098  106   122   126  120   128 172  118  054   212   176  104   080  156 242  122
 248  220   172   222  078   042 204  046 158    032   030  210   058  174 164  206
 222  076   154   216  216   094 102  032 030    238   156  246   126  144 252  134
 120  236   182   214  050   156 022  072 248    032   234  072  222   188  228  121
```

In this we have excluded the ciphertext which is already presented in (3.8)

In the light of this analysis, here we conclude that this cipher is an interesting one and a strong one, and this can be used for the transmission of any information through internet.

References

[1] V.U.K Sastry and K. Anup Kumar “ A Novel Feistel Cipher Involving a bunch of Keys Supplemented with XOR Operation” (IJACSA) International Journal of Advanced Computer Science and Applications, 2012.

[2] William Stallings, Cryptography and Network Security, Principles and Practice, Third Edition, Pearson, 2003.

Authors Profile

Dr. V. U. K. Sastry is presently working as Professor in the Dept. of Computer Science and Engineering (CSE), Director (SCSI), Dean (R & D), SreeNidhi Institute of Science and Technology (SNIST), Hyderabad, India. He was Formerly Professor in IIT, Kharagpur, India andWorked in IIT, Kharagpurduring 1963 – 1998. He guided 12 PhDs, and published more than 40 research papers in various international journals. His research interests are Network Security & Cryptography, Image Processing, Data Mining and Genetic Algorithms.

Mr. K. Anup Kumar is presently working as an Associate Professor in the Department of Computer Science and Engineering, SNIST, Hyderabad India. He obtained his B.Tech (CSE) degree from JNTU Hyderabad and his M.Tech (CSE) from Osmania university, Hyderabad. He is now pursuing his PhD from JNTU, Hyderabad, India, under the supervision of Dr. V.U.K. Sastry in the area of Information Security and Cryptography. He has 10 years of teaching experience and his interest in research area includes, Cryptography, Steganography and Parallel Processing Systems.

A Novel Feistel Cipher Involving a Bunch of Keys Supplemented with XOR Operation

V.U.K Sastry
Dean R&D, Department of Computer Science and Engineering, Sreenidhi Institute of Science & Tech.
Hyderabad, India

K. Anup Kumar
Associate Professor, Department of Computer Science and Engineering, SNIST,
Hyderabad, India

***Abstract*—In this investigation, we have developed a novel block cipher by modifying classical Feistel cipher. In this, we have used a key bunched wherein each key has a multiplicative inverse. The cryptanalysis carried out in this investigation clearly shows that this cipher cannot be broken by any attack.**

Keywords-encryption; decryption; cryptanalysis; avalanche effect; multiplicative inverse.

I. Introduction

The study of the Feistel cipher [1-2] laid the foundation for the development of cryptography in the seventies of the last century. In the classical Feistel cipher the block size is 64 bits, and it is divided into two halves wherein each half is containing 32 bits. The number of rounds in the iteration process is 16. The basic equations governing the Feistel cipher can be written in the form

$$P^i = Q^{i-1}, \qquad (1.1)$$
$$Q^i = P^{i-1} \oplus F(Q^{i-1}, K^i), \qquad (1.2)$$

and

$$Q^{i-1} = P^i, \qquad (1.3)$$
$$P^{i-1} = Q^i \oplus F(P^i, K^i), \qquad (1.4)$$

where P^i and Q^i are the blocks of the plaintext in the ith round of the iteration process, F is a function chosen appropriately, and K^i is the key in the ith round. In this analysis, the XOR operation and the permutation that is performed by interchanging two halves of the data in the iteration process play a vital role in deciding the strength of the cipher.

In the recent years, Sastry et al. [3-12] have offered several modifications to the Feistel cipher, and have studied various aspects of this cipher, including different types of permutations and substitutions. In all these investigations we have divided the plaintext into a pair of matrices of equal size, and the key is taken in the form of a matrix.

In the process of encryption, we take the key bunch as E, and represent it in the form of a matrix given by $E = [e_{jk}]$. The corresponding key bunch in the process of decryption is taken as $D = [d_{jk}]$. Here for a given value of the key e_{jk}, used in the encryption, we determine the corresponding key d_{jk}, by using the relation

$$(e_{jk} \times d_{jk}) \bmod 256 = 1, \qquad (1.5)$$

where d_{kl} is the multiplicative inverse of e_{kl},

In order to satisfy (1.5), we chose e_{jk} as an odd integer, which lies in the interval [1-255], and thus we obtain d_{jk} also as an odd integer lying in the interval [1-255].

Here also we adopt an iterative procedure, and make use of the permutation process that consists of the interchange of the two halves of the plaintext , of course, represented in the form of a pair of matrices.

In the present analysis, our objective is to modify the Feistel cipher by including a bunch of keys. Here our interest is to see how the different keys, occurring in the key bunch, would influence the strength of the cipher.

In what follows, we present the plan of the paper. In section 2, we introduce the development of the cipher and present the flowcharts and the algorithms corresponding to the cipher, in section 3, we illustrate the cipher with an example and examine the avalanche effect. After that, we carry out the cryptanalysis in section 4. Finally we present numerical computation and draw conclusions.

II. Development Of The Cipher

We consider a plaintext P containing $2m^2$ characters. On using the EBCIDIC code this is written in terms of numbers which are in the interval [0-255]. Now we write this in the form of a pair of square matrices P^0 and Q^0, wherein each matrix is of size m.

The basic equations governing the encryption of this block cipher are given by

$$[P_{jk}^{\ i}] = [e_{jk}\ Q_{jk}^{\ i-1}] \bmod 256, \qquad (2.1)$$

and

$$[Q_{jk}^{\ i}] = [e_{jk}\ P_{jk}^{\ i-1}] \bmod 256 \oplus [Q_{jk}^{\ i-1}], \qquad (2.2)$$

where j= 1 to m, k = 1 to m and i =1 to n, in which n is the number of rounds.

The corresponding equations of decryption are in the form,

$$[Q_{jk}^{i-1}] = [d_{jk} P_{jk}^{i}] \bmod 256, \qquad (2.3)$$

and

$$[P_{jk}^{\ i-1}] = [d_{jk}([Q_{jk}^{\ i}] \oplus [Q_{jk}^{\ i-1}])] \bmod 256 \qquad (2.4)$$

where j= 1 to m, k = 1 to m and i = n to 1, here P_{jk}^i and Q_{jk}^i are the jth row kth column elements of the left and right portions of the plaintext matrix, respectively, in the ith round of the iteration process.

On using the basic relations (2.1) - (2.4), governing the encryption and the decryption, the corresponding flowcharts for the encryption and the decryption can be written as shown below.

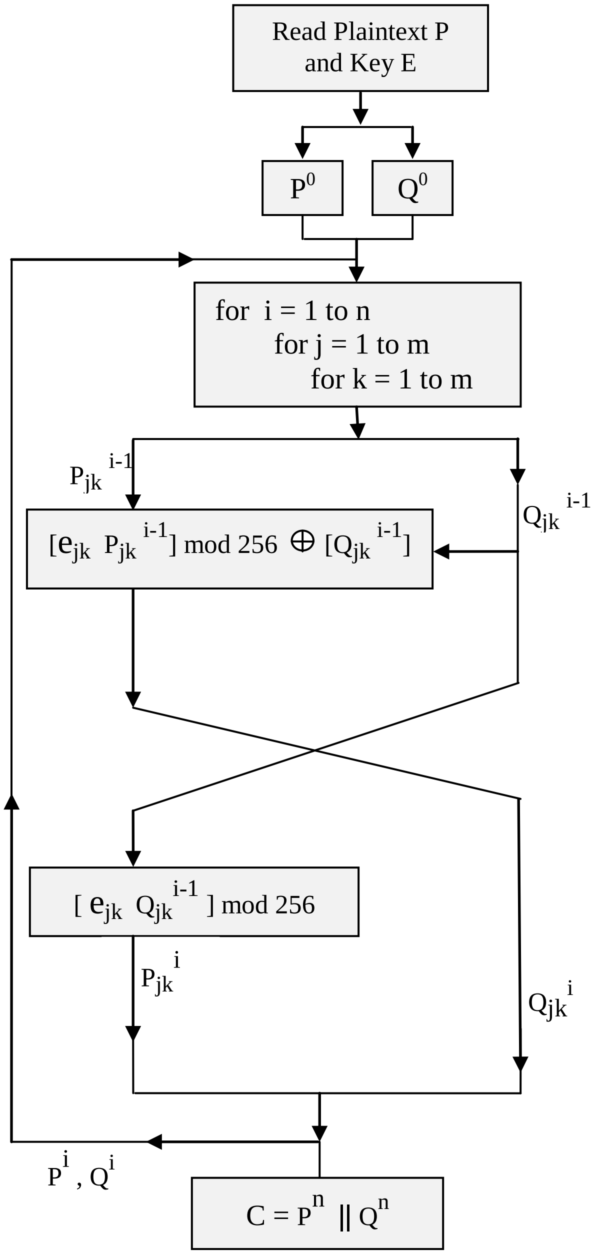


Figure 1. The Process of Encryption

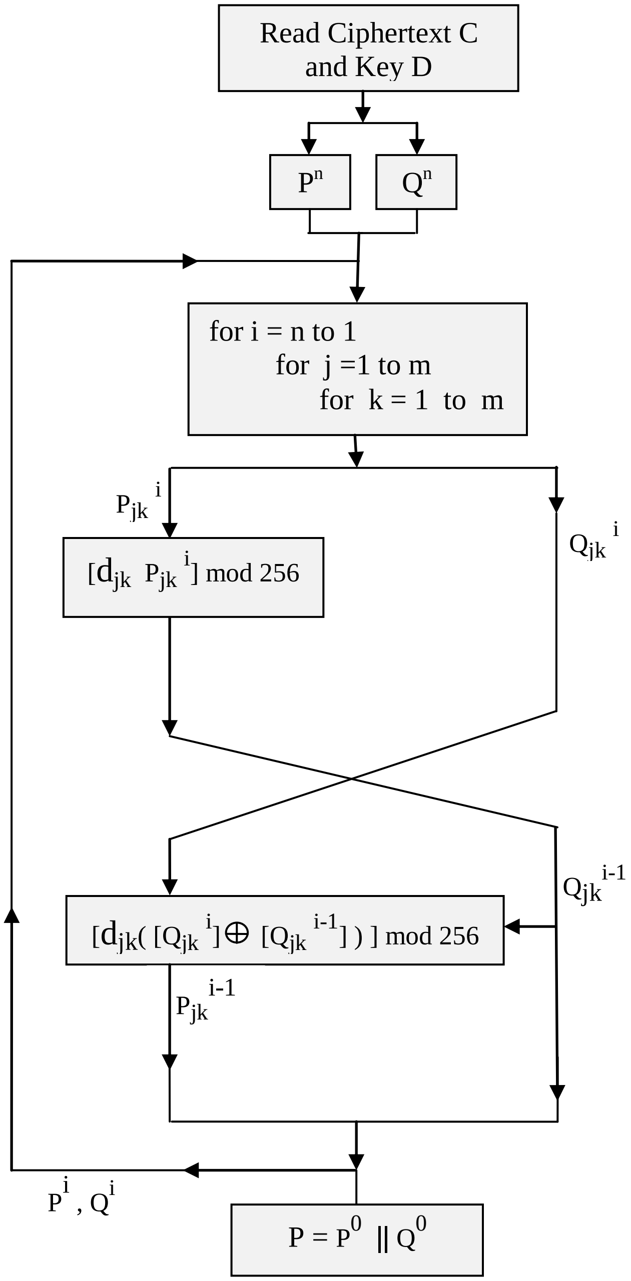


Figure 2. The process of Decryption

The algorithms for the encryption and the decryption are written as shown below.

A. Algorithm for Encryption

```
1. Read P, E, and n
2. P^0 = Left half of P.
   Q^0 = Right half of P.
3. for i = 1 to n
```

```
begin
    for j = 1 to m
    begin
        for k = 1 to m
            begin
              $[ P_{jk}^{i} ]= [ e_{jk} Q_{jk}^{i-1}] \bmod 256$,
             $[ Q_{jk}^{i} ]= [e_{jk} P_{jk}^{i-1}] \bmod 256 \oplus [Q_{jk}^{i-1}]$ ,
            end
     end
  end
6. $C = P^n \| Q^n \|$ /* represents concatenation */
7. Write(C)
```

B. Algorithm for Decryption

```
1. Read C, D, and n.
2. $P^n$ = Left half of C
   $Q^n$ = Right half of C
3. for i = n to 1
begin
    for j = 1 to m
      begin
          for k = 1 to m
            begin
            $[Q_{jk}^{i-1}] = [ d_{jk} P_{jk}^{i} ] \bmod 256$,

          $[P_{jk}^{i-1}]=[d_{jk} ([Q_{jk}^{i}] \oplus [Q_{jk}^{i-1}]] \bmod 256$
           end
     end
end
6. $P = P^0 \| Q^0$ /*$\|$ represents concatenation */
7. Write (P)
```

In what follows we illustrate the cipher with a suitable example.

III. Illustration Of The Cipher

Consider the plaintext given below.

Brother! When we were very poor, by looking at some corrupt politicians and employees who earned crores and crores, we use to think how to come up in life. Though you were a graduate having technical skills, you joined naxalites thinking that only unethical rich are totally responsible for the ruination of our country. After the death of our father, I joined as a police, our uncle started liquor business! He has earned crores and crores. Though I have become a police inspector, I am helpless. I am not able to control anything! I do not know when India will change! Write a letter. Do come back. (3.1)

Let us focus our attention on the first 32 characters of the plaintext (3.1). This is given by

Brother! When we were very poor, (3.2)

On using EBCIDIC code, we write (3.2) in the form of a pair of square matrices given by

$$P^0 = \begin{bmatrix} 066 & 114 & 111 & 116 \\ 032 & 087 & 104 & 101 \\ 032 & 119 & 101 & 114 \\ 114 & 121 & 032 & 112 \end{bmatrix} \quad (3.3)$$

and

$$Q^0 = \begin{bmatrix} 104 & 101 & 114 & 033 \\ 110 & 032 & 119 & 101 \\ 101 & 032 & 118 & 101 \\ 111 & 111 & 114 & 044 \end{bmatrix} \quad (3.4)$$

Let us take the encryption key bunch matrix E in the form

$$E = \begin{bmatrix} 071 & 053 & 011 & 061 \\ 117 & 069 & 057 & 051 \\ 121 & 139 & 101 & 043 \\ 099 & 095 & 111 & 035 \end{bmatrix} \quad (3.5)$$

and

$$D = \begin{bmatrix} 119 & 029 & 163 & 021 \\ 221 & 141 & 009 & 251 \\ 201 & 035 & 109 & 131 \\ 075 & 159 & 143 & 139 \end{bmatrix} \quad (3.6)$$

On using the algorithm for encryption, given in section 2, we get the ciphertext C in the form

$$C = \begin{bmatrix} 088 & 166 & 064 & 218 & 222 & 060 & 064 & 088 \\ 140 & 078 & 014 & 104 & 028 & 204 & 176 & 036 \\ 088 & 094 & 002 & 182 & 244 & 188 & 202 & 108 \\ 240 & 120 & 038 & 118 & 208 & 224 & 146 & 196 \end{bmatrix} \quad (3.7)$$

On using the keys in D, and applying the decryption algorithm on (3.6), we get back the original plaintext P given by (3.2)

Let us now study the avalanche effect. On changing the first row , second column element of P_0 from 114 to 115, we get a change of one binary bit in the plaintext.

On applying the encryption algorithm on this modified plaintext, using the same key bunch matrix E, we get the ciphertext C in the form

$$C = \begin{bmatrix} 104 & 144 & 028 & 204 & 176 & 122 & 222 & 228 \\ 172 & 244 & 236 & 162 & 210 & 024 & 158 & 030 \\ 214 & 206 & 016 & 004 & 144 & 096 & 120 & 014 \\ 218 & 218 & 226 & 242 & 076 & 036 & 176 & 086 \end{bmatrix} \quad (3.8)$$

On converting (3.6) and (3.8) into their binary form and comparing them, we notice that these two ciphertext differ by 133 bits out of 256 bits. This shows that the strength of the cipher is quite up to the mark.

Now let us consider one bit change in the key bunch matrix E. On replacing first row , second column element of E from 53 to 52, we have a one bit change in the key matrix. On using the original plaintext (3.2) and the modified key bunch matrix, and the algorithm for encryption, given in section 2, we get the ciphertext C in the form

$$C = \begin{bmatrix} 246 & 150 & 038 & 080 & 058 & 246 & 202 & 246 \\ 190 & 170 & 220 & 124 & 038 & 238 & 178 & 202 \\ 230 & 040 & 236 & 250 & 004 & 036 & 154 & 022 \\ 224 & 122 & 166 & 216 & 146 & 218 & 182 & 238 \end{bmatrix} \quad (3.9)$$

On comparing (3.6) and (3.9), in their binary form, we find that two ciphertext matrices differ by 127 bits out of 256 bits. This also shows that the block cipher under consideration is a potential one.

IV. Cryptanalysis

The different types of cryptanalytic attacks that are well known in the literature of cryptography [13] are

1. Ciphertext only attack (Brute Force Attack),
2. Known plaintext attack,
3. Chosen plaintext attack, and
4. Chosen ciphertext attack.

Generally all the algorithms are designed to withstand the brute force attack and the known plaintext attack. Further, an algorithm is examined, intuitively, whether it withstands the later two attacks or not.

In this analysis, the key bunch is a square matrix containing m2 elements. In view of this fact, the size of the key space is

$$2^{(8m^2)} = (2^{10})^{0.8\,m^2} \approx (10^3)^{0.8\,m^2} = (10)^{2.4m^2}$$

If we assume that the time required for encryption with each key in the key space as 10-7 seconds, then the time required for all the keys in the key space is approximately equal to

$$\frac{10^{(2.4m^2)} \times 10^{-7}}{365 \times 24 \times 60 \times 60} \text{ years} = 3.12 \times 10^{(2.4\,m^2 - 15)} \text{ years}$$

In this analysis, we have taken m=4. Thus the time required is 3.12 x 10 23.4 years.

As this time is very large, we cannot break the cipher by the brute force attack.

Let us now consider the known plaintext attack. If we confine our attention only to one round of the iteration process, that is when n=1, from the encryption algorithm given in section 2, we get

$$[\, P_{jk}^{1} \,] = [\, e_{jk}\, Q_{jk}^{0} \,] \bmod 256, \quad (4.1)$$

and

$$[\, Q_{jk}^{1} \,] = [\, e_{jk}\, P_{jk}^{0} \,] \bmod 256 \oplus [\, Q_{jk}^{0} \,], \quad (4.2)$$

where, j = 1 to m and k = 1 to m.

Further we have,

$$C = P^{1} \parallel Q^{1} . \quad (4.3)$$

In the known plaintext attack, the attacker knows the plaintext and the corresponding ciphertext. Thus he knows P_{jk}^{1} and Q_{jk}^{1} occurring in (4.3) and, P_{jk}^{0} and Q_{jk}^{0} occurring in (4.1) and (4.2). in the light of this fact on obtaining the multiplicative inverse of Q_{jk}^{0} (selecting Q_{jk}^{0} as odd numbers) occurring in (4.1), the attacker can determine e_{jk} very conveniently. Thus the cipher can be broken when n=1.

Let us known consider the case when n=2. In this case, the equations governing the encryption are of the form

$$[\, P_{jk}^{1} \,] = [\, e_{jk}\, Q_{jk}^{0} \,] \bmod 256, \quad (4.4)$$

and

$$[\, Q_{jk}^{1} \,] = [\, e_{jk}\, P_{jk}^{0} \,] \bmod 256 \oplus [\, Q_{jk}^{0} \,], \quad (4.5)$$

$$[\, P_{jk}^{2} \,] = [\, e_{jk}\, Q_{jk}^{1} \,] \bmod 256, \quad (4.6)$$

and

$$[\, Q_{jk}^{2} \,] = [\, e_{jk}\, P_{jk}^{1} \,] \bmod 256 \oplus [\, Q_{jk}^{1} \,], \quad (4.7)$$

where, j = 1 to m and k = 1 to m.

Further we have,

$$C = P^{2} \parallel Q^{2} . \quad (4.8)$$

As C is known, the attacker knows P_{jk}^{2} and Q_{jk}^{2}. The attacker also knows P_{jk}^{0} and Q_{jk}^{0} occurring in (4.4) and (4.5) as this is the known plaintext attack. Thus the attacker cannot determine the keys e_{jk} occurring in (4.4) as P_{jk}^{1} cannot be determined by any means.

In the light of aforementioned fact, this cipher having sixteen rounds (n=16) cannot be broken by the known plaintext attack.

On inspecting the above equations arising in this analysis, we find that it is simply impossible to choose plaintext or ciphertext, intuitively, and break the cipher in any way. Thus we conclude that this cipher is a strong one.

V. Computations And Conclusions

In this investigation, we have developed a block cipher by selecting a bunch of keys wherein each key has its multiplicative inverse.

The programs for the encryption and the decryption are written in C language.

Let us now consider the entire plaintext given by (3.1). This can be divided into 19 blocks wherein each block is having 32 characters. As the last block is containing 24 characters, we append 8 blank characters. On using the key bunch E, given by (3.5), and carrying out the encryption process given in section 2, we get the ciphertext (excluding the portion which is already given by (3.6)), we get

```
 142  090  154  198  076 066  218  154  144  090  026  248  178  024  218  182
 038  250  088  006  110 124  240  000  102  048  180  188  172  118  054  212
 176 104   080  156  242 070  214  198  228  102  250  092  228  190  250  074
 020 102   152  006  110 076  098  106  122  126  120  128  172  118  054  212
 176 104   080  156  242 070  214  198  228  102  250  092  228  190  250  074
 020 102   150  206  016 014  104  028  214  098  018  194  110  094  150  150
 048 236   170  216  012 158  142  228  194  134  076  242  072  098  172  210
 032 236   224  134  056 110  246  036  254  044  244  054  204  144  110  120
 206 222   082  230  238 166  204  236  236  174  178  016  014  118  078  250
 062 072   126  066  188 246  218  232  002  200  150  036  242   054 038  116
 042 232   144  052  048 140  114  098  174  134  114  110  130  102  036  142
 012  068  004  236  232 114  082  086  138  098  072  140  182  192  218  230
 112  246  122  220  156 188  006  106  088  200  218  070  156  182  050  156
 022  072  248  034  232 172  090  036  172  092  122  210  118  238  056  222
 230 044   254  210  094 076  098  216  036  098  236  238  072  254  090  234
 108 172   022  198  146 028  182  054  140  154  134  182  054  034  182  054
 240 102   048  180  110 076  244  178  014  222  248  226  002  204  098  106
 122 090   236  108  170 098  210  162  056  228  142  174  140  202  204  244
 184 202   126  244  150 042  204  050  014  222  152  198  214  246  252  240
 000 090   236  108  170 098  210  162  056  228  142  174  140  202  204  244
 226 172   210  248  226 000  236  034  018  204  098  120  108  202  242  210
 198 028  180   090  000 178  208  220  152  112  232  158  104  044  084  154
 100 028  188   016  230 044  204  144  110  120  206  222  082  230  238  166
 204 236  236   174  178 016  014  118  078  250  062  072  126  066  188  246
 218 232  002   200  150 036  242  054  038  116  042  232  144  052  048  140
 114 0 98 174   134  114 110  130  102  036  142  012  068  004  236  232  114
 082 098  044   176  118 248  156  060  158  182  038  110  000  218  150  050
 118 208  230   108  140 230  004  146  062  072  122  210  118  238  056  032
 024 140  078   012  184 126  090  088  194  182  170  096  054  122  058  146
 014 028  050   204  036 138  178  076  130  182  130  028  228  184  146  044
 238 056  250   176  224 136  128  188  188  046  074  076  100  182  014  222
 050 134  178   214  228 230  044  254  210  094  076  098  216  036  098  236
 238 072  254   090  234 108  172  022  198  146  028  182  054  140  154  134
 182 054  034   182  054 240  102  048  180  110  076  244  178  014  222  248
 226 002  204   098  106 122  090  236  108  170  098  210  162  056  228  142
 174 140  202   204  244 184  202  126  244  150  042  204  050  014  222  152
 198 214  246   252  240 000  090  236  108  170  098  210  162  056  228  244
 240 184  088   190  156 084  154  094  060  064  060  164  118  092  074  158
```

From the above analysis we conclude that the novel Feistel cipher, wherein we have made use of a bunch of keys is a strong one as the cryptanalysis shows that it cannot be broken by any attack. This is all on account of the iteration process and the multiplication by the bunch of keys.

References

[1] William Stallings, Cryptography and Network Security, Principles and Practice, Third Edition, Pearson, 2003.

[2] Feistel H, "Cryptography and Computer Privacy", Scientific American, Vol. 228, No.5, pp. 15-23, 1973.

[3] V.U.K. Sastry and K. Anup Kumar, "A Modified Feistel Cipher involving a key as a multiplicant on both the sides of the Plaintext matrix and supplemented with Mixing, Permutation and XOR Operation", International Journal of Computer Technology and Applications, ISSN 2229-6093,Vol 3 (1), pp, 23-31 , 2012.

[4] V.U.K. Sastry and K. Anup Kumar, "A Modified Feistel Cipher involving a key as a multiplicant on both the sides of the Plaintext matrix and supplemented with Mixing, Permutation and Modular Arithmetic Addition", International Journal of Computer Technology and Applications, ISSN 2229-6093,Vol 3 (1), pp, 32-39 , 2012.

[5] V.U.K. Sastry and K. Anup Kumar, "A Modified Feistel Cipher Involving a Key as a Multiplicant on Both the Sides of the Plaintext Matrix and Supplemented with Mixing, Permutation, and Modular Arithmetic Addition", International Journal of Computer Science and Information Technologies ISSN 0975 – 9646. Vol. 3 (1) , 2012, pp, 3133 – 3141,2012.

[6] V.U.K. Sastry and K. Anup Kumar, "A Modified Feistel Cipher involving a pair of key matrices, Supplemented with Modular Arithmetic Addition and Shuffling of the plaintext in each round of the iteration process", International Journal of Computer Science and Information Technologies ISSN 0975 – 9646. Vol. 3 (1) , 2012, pp, 3119 – 3128,2012.

[7] V.U.K Sastry and K. Anup Kumar, "A Modified Feistel Cipher Involving XOR Operation and Modular Arithmetic Inverse of a Key Matrix" International Journal of Advanced Computer Science and Applications ISSN : 2156-5570(Online), pp.35-39 , Vol.3 No.7. 2012, U.S.A

[8] V.U.K Sastry and K. Anup Kumar, "A Modified Feistel Cipher Involving Modular Arithmetic Addition and Modular Arithmetic Inverse of a Key Matrix" International Journal of Advanced Computer Science and Applications ISSN : 2156-5570(Online),pp. 40-43, Vol.3 No.7, 2012, U.S.A.

[9] V.U.K Sastry and K. Anup Kumar, "A Modified Feistel Cipher Involving Substitution, Shifting of rows, Mixing of columns, XOR operation with a Key and Shuffling" International Journal of Advanced Computer Science and Applications ISSN : 2156-5570(Online), pp. 23-29, Vol.3 No.8, 2012, U.S.A.

[10] V.U.K Sastry and K. Anup Kumar, "A Modified Feistel Cipher Involving Key Based Substitution, Shifting of rows, Key Based Mixing of columns, Modular Arithmetic Addition and Shuffling "International Journal of Engineering Research and Applications (IJERA) ISSN : 2248-9622 (Online), Vol. 2 , No. 5, pp 237-245, 2012.

[11] V.U.K Sastry and K. Anup Kumar, "A Block Cipher Obtained by Blending Modified Feistel Cipher and Advanced Hill Cipher Involving a Single Key Matrix International Journal of Engineering Research and Applications (IJERA) ISSN : 2248-9622 (Online), Vol. 2 , No. 5, pp 951-958, 2012.

[12] V.U.K Sastry and K. Anup Kumar, "A Block Cipher Obtained by Blending Modified Feistel Cipher and Advanced Hill Cipher Involving a Pair of Key Matrices International Journal of Engineering Research and Applications (IJERA) ISSN : 2248-9622 (Online), Vol. 2 , No. 5, pp 959-964, 2012.

Authors Profile

Dr. V. U. K. Sastry is presently working as Professor in the Dept. of Computer Science and Engineering (CSE), Director (SCSI), Dean (R & D), SreeNidhi Institute of Science and Technology (SNIST), Hyderabad, India. He was Formerly Professor in IIT, Kharagpur, India andWorked in IIT, Kharagpurduring 1963 – 1998. He guided 12 PhDs, and published more than 40 research papers in various international journals. His research interests are Network Security & Cryptography, Image Processing, Data Mining and Genetic Algorithms.

Mr. K. Anup Kumar is presently working as an Associate Professor in the Department of Computer Science and Engineering, SNIST, Hyderabad India. He obtained his B.Tech (CSE) degree from JNTU Hyderabad and his M.Tech (CSE) from Osmania university, Hyderabad. He is now pursuing his PhD from JNTU, Hyderabad, India, under the supervision of Dr. V.U.K. Sastry in the area of Information Security and Cryptography. He has 10 years of teaching experience and his interest in research area includes, Cryptography, Steganography and Parallel Processing Systems

Computing the Exit Complexity of Knowledge in Distributed Quantum Computers

M.A.Abbas
Electrical Engineering Department
Faculty of Engineering
King Khaled University, KSA

Abstract— **Distributed Quantum computers abide from the exit complexity of the knowledge. The exit complexity is the accrue of the nodal information needed to clarify the total egress system with deference to a distinguished exit node. The core objective of this paper is to compile an arrogant methodology for assessing the exit complexity of the knowledge in distributed quantum computers. The proposed methodology is based on contouring the knowledge using the unlabeled binary trees, hence building an benchmarked and a computer based model. The proposed methodology dramatizes knowledge autocratically calculates the exit complexity. The methodology consists of several amphitheaters, starting with detecting the baron aspect of the tree of others entitled express knowledge and then measure the volume of information and the complexity of behavior destining from the bargain of information. Then calculate egress resulting from episodes that do not lead to the withdrawal of the information. In the end is calculated total egress complexity and then appraised total exit complexity of the system. Given the complexity of the operations within the Distributed Computing Quantity, this research addresses effective transactions that could affect the three-dimensional behavior of knowledge. The results materialized that the best affair where total exit complexity as minimal as possible is a picture of a binary tree is entitled at the rate of positive and negative cardinal points medium value. It could be argued that these cardinal points should not amass the upper bound apex or minimum**.

Keywords-Complexity; Quantum Computers; Knowledge acquisition; Graph theory; Egress; Distributed systems.

I. Introduction

Computing the degree of complexity is a problem that faces the knowledge in Distributed Computing and working quantum theory and is considered one of the most important obstacles facing computer scientists during the study of this type of computing. In prevalent, the behaviors are appeared on quantum computing complicated activities to the bounds apprise. The quantum computers are appraisal appliances that cultivate govern function of quantum mechanical mahatma, such as superposition and bafflement, to comport activities on evidence. Quantum computers are asymmetric with integral computers based on transistors. Whereas integral computers need evidence to be encoded into bilateral characters bits, quantum computation eases quantum chattels to approximate data and comport tries on these data [1]. We can certify the dimension of the eminence of knowledge in this brand of computers, precisely when chattering about the broken up assessing and the eagerness for consonance between the mobile data as beneficially as information brought bounteous computers affixed the knowledge ensuing from this information. In a distributed system, creations dispatch with exhaustive ambiences passing in a choosy protocol, and potential universes are computed by a sameness analogy above these aspects. absolutely, all crossway aspects are approximated indistinguishable adjacent an appointee if its close-at-hand aspect in these circumstances are identical. eloquently, one has to analyze concernedly this when concentrating distributed networks where quantum assets are apportioned. [2-24]

The Exit complexity is the accumulate of the nodal caution needed to explicate a total egress system with esteem to a specialized exit node. Egress is the behave of leaving out of matter. For archetypal, in telecommunications, an egress router is a router intruding which a datum packet departs one network for another network. As for quantum information, an agent perceives which qubits it claims, what close-at-hand activities it conducts on these qubits, along with, as well as, what non-local beginning aspect it buds away with what bafflement it associates with disjoint alternates primarily. It can additionally acquire the details on its adjacent quantum enters, furthermore this is not obligatorily. An agent does not requisitely discern anything about its qubits, they may be unknown enters, but they may also have been pocketed as a quantum post from another appointee. This is why the abridged compactness matrix is not a better auger to clarify quantum knowledge of single alternates: it ascribes too much information for agents monopolizing qubits they perceive nothing about, and together with concise information for alternates formatively allocating a grasped bafflement drafting. [2-24]

This paper advances the rule-based modeling of the exit complexity of knowledge in distributed quantum computers. The underlying philosophy stems from the belief that knowledge in distributed quantum computers may be under primed by classical graph theory. The main objective of this paper is to analyze and compute the exit complexity of knowledge during it's running off from distributed quantum computers. This analysis ensures that the knowledge complies with the system requirements and conditions. Next sections are organized as follow: section two presents the previous works in this field and the formal knowledge representation in distributed systems. Section three introduces the proposed knowledge representation in distributed quantum computers. Section four represents a simulation environment of quantum computer systems and the results of the proposed models and algorithms. The results focus on the computing of the exit complexity of the knowledge.

II. Related Works

This section clarifies the bygone effort in quantum computer appraisals. The conception of quantum approximation In the began 1990s bounteous crafts persons [25] accosted algebraic behaviors which could be disinterred by a quantum computer accrual conveniently than any realistic computer. Such a quantum algorithm would amusement a dreamy role analogous to that of Bell’s asymmetry, in denoting object of the extant nature of quantum mechanics. primarily loan very bantam alterations in behavior were disclosed, in which quantum mechanics affirmed a comeback to be found with conviction, hence the quantum system was noise-free, where a probabilistic constructivist computer could apprehend a comeback ‘only’ with atmospheric liability. A considerable approach was brought by Simon[26], who clarified a convenient quantum approximation for dilemma for which no convenient breakthrough was available decorously, gradual adjacent probabilistic appraisals. This enlivened Shor [27] who bewildered the community by explicating an appraisal which was not only convenient on a QC, furthermore also approached a core dilemma in computer science: that of factorizing bulky figures. conceptions of quantum information and quantum considering converge together. For, a QC can be made much beneath attentive to noise by a beginning belief which closes eloquently from the envelopment of quantum mechanics with constructivist information perception, namely quantum inaccuracy correction. furthermore the phraseology ‘error counteraction’ is a broad one and was exerted with it was individual in that decade that two big-name papers, and connected an accustomed configuration whereby quantum information processing barrages a very broad class of acoustics manners in a conscientiously contemplated quantum computation. Much blessing has hence been caused in implying these beliefs. A consequential conception was the exhibit by Kitaev [28] that counteraction can be apprehended continual when the healing behaviors are themselves blotched. imitative computations direct to an omnipresent design of ‘fault merciful’ counting, of which an auxiliary observe is ascribed by Preskill [29]. Their claim been numerous approaches at explicating quantum schemas in provisos of formulation transformers: in [30] a synthesis of quantum assesses, based on propositional energetic deduction is constructed, while in [31], a weakest precondition semantics for quantum considerations are corrected. additionally, a propositional cogitation conceived to explicate quantum determination at an operational category is embed obtrusive in [32]. These architectures, albeit, direction at cloning condoned counts that audit an input-output partnership, a forepart of spectacle which is not acceptable for allocated evaluations. A beginning approach to denote knowledge for quantum assigned systems is found in [32]. There, the allowed application is impended, i.e., a protocol is aped as a budget of melts and an equality involvement is denoted for each deputy alternate on connotations of these melts. Two concepts of knowledge, one constructivist and one quantum, are inferred, based on differing designs of balance. Egress Complexity is an approach distinct, non-metric methodology that apprises the egress ability of an apportioned environment. The assign of adeptness is commandingly the egress route complexity – formatively developed to distribute the anticipation continued by a naïve incumbent in aim of an exit [33]. additional currently, the amplitude of the methodology for apprising strive complexity has been determined by the authors. Every conceiving has a concealed measure of coarse complexity that may be accounted about from the circumstance of the creating’s forum concept. The amplitude of this complexity is catatonic criterion of the alike topological network of dormitories and affixing colonnades and is hence a generating different apportion quite conflicting from forecast based assignments of time-required modeling.

A. Formal Knowledge Representation in Distributed Systems

This subsection convinces the anteceding work that apes the Knowledge in distributed systems. analyzing knowledge events in a distributed system is begun in [2], they clarified formally how knowledge evolves as the algebra behaves. knowledge can be circumscribed in distributed systems multi-instance bi-form trees reciprocal canisters. In this analyze, we assign these canisters tree in Figure 1. The trees are co-active knowledge approximated allowed the multiplicity of their circumstances in quantum computers. Multi-instance binary trees are tree data formulations in which each node has at most two descendant nodes. Nodes with children are creator nodes, and descendant nodes may compose mentions to their parents. alien the tree, there is frequent a quotation to the root node. numerous node in the data configuration can be reached by beginning at root node and recurrently consequent mentions to matching the left or right descendant. This can be behaved plentifully by brewing knowledge operators Ki with the timed operators denoted in the former section. So, they probe how knowledge evolves in distributed systems due to message coursing. To work out what a broadcasting such as apparently, easing an amalgamate of alert from lengthwise timed logic. The most permissible aspects that approximate the knowledge is cheered affirmed synchronous aspect [2]. The article timed logic is used to clarify any system of masters and symbolism for casting, and cogitating about, anticipations exceptional in names of time. They introduce the traditional temporal state operators , always , and , eventually, combined with the path operators A for all paths and E there exists multiple paths , as follows : the path , the path , the path and the path all of them are depicted in figure1.

On multi-instance tree awareness, we compactly acquire additionally multi-instance clause we expect for this discussion. We identify a purse positive or negative according to its allotted brand. As the data set does not enclose the true caste of all lone instances, we call an instance positive if it is part of a positive bag. We call a positive example a true

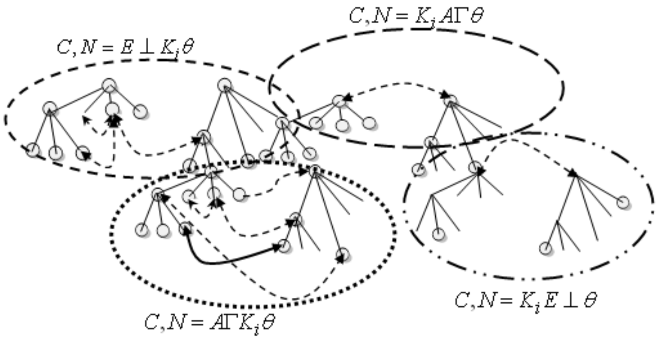

$C,N = E \perp K_j\theta$
$C,N = K_j A\Gamma\theta$
$C,N = A\Gamma K_i\theta$
$C,N = K_i E \perp \theta$

Fig. 1 A schematic representation of multi-instance binary trees [2]

positive, if its true class is positive; and a false positive, if its true class is negative. All instances in negative bags are called negative. Consider PI represents the total of positive instances and NI the total of negative instances. Using the pseudo-information measure suggested in [34, 35], the following equation calculates the information needed to egress from any no exit node

$$I = -PI\log_2\frac{PI}{PI+NI} - NI\log_2\frac{NI}{PI+NI} \quad (1)\ \ [34\,]$$

Hence, E(S) relative to e ~ is given by in the system to e':

$$E(s) = \sum_k I = -\sum_k PI\log_2\frac{PI}{PI+NI} - NI\log_2\frac{NI}{PI+NI} \quad (2)\ [34\,]$$

The number rn of unlabeled rooted trees is given by the formula

$$r_n = \sum_{t1+2j2+.....+(n-1)tn-1=n-1} \prod_{k=1}^{n-1}\binom{r_k + j_k - 1}{t_k} \quad (3)\ \ [35$$
]

The number of positive instances PI always exceeds the number of negative instances NI. Since the path to the exit node can be traversed at most once.

III. PROPOSED KNOWLEDGE REPRESENTATION IN DISTRIBUTED QUANTUM COMPUTERS

This section presents the proposed representation of the knowledge hierarchy in distributed quantum computers. As free trees that are a connected graph without cycles. The number rn of unlabeled rooted trees is inherited from equation 3 as follow:

$$r_n = PI + NI = \sum_x \prod_{d=1}^{h-1}\binom{r_d + j_d - 1}{m_d}$$

Where $x = m1 + 2m2 + ... + (h-1)mh - 1 = h - 1$

In this case, this paper assumes the number of unlabeled roots trees is the sum of both positive instances and negative instances.

After computing the total number of instances. We can compute the exit complexity as depicted in the formula in equation4

$$E(s) = \sum_k I = -\sum_k PI\log_2\frac{PI}{r_n} - n^{-}\log_2\frac{NI}{r_n} \quad (4)$$

To calculate the Kinetic complexity of the unlabeled tree, we use main equations described in [36]. The Kinetic equation starts with the following formula, where the total exit complexity is an aggregation of both kinetic complexity and the egress complexity.

$$\frac{\partial k}{\partial t}(v,t) = \int k(v1,t)K(v) - k(v,t)K(v1))\sigma(v,v1)dv1 \quad (5)$$

The kinetic equations models the velocities of the knowledge transfer, with velocities v introduced at t=0. And time is divided into time-steps of duration Δt. In N spatial dimensions the site index n is replaced by a vector and the quadratic approximation to the kinetic operator T becomes

$$(H\lambda) = \frac{O^2}{2sa^2}\left[2U\lambda_n - \sum_{a=1}^{U}(\lambda_{n-a} + \lambda_{n+a})\right] \quad (6)$$

Where O is Hamiltoman operator , s is the mass of a single particle ,λ represents an eigenvalue, Hλ is the kinetic operator , and a is a uniform distance spacing. After determining the kinetic operator, the kinetic complexity could be represented from equations 5 and 6 as depicted in the following formulas in equation 7.

$$(H\lambda)\times\frac{\partial k}{\partial t}(v,t) = \left(\bar{T}\,\lambda\right)\times\left[\int k(v1,t)K(v) - k(v,t)K(v1))\sigma(v,v1)dv1\right] \quad (7)$$

At this point, the total exit complexity should be realized as follow: Total Exit complexity (TEC) is the sum of both Egress Complexity and Kinetic Complexity. More expands of this equation appear by merging equation 4 and 7. then (TEC) should be represented as follow:

$$TEC = (H\lambda)\times\left[\int k(v1,t)K(v) - k(v,t)K(v1))\sigma(v,v1)dv1\right] - \sum_k PI\log_2\frac{PI}{r_n} - NI\log_2\frac{NI}{r_n} \quad (8)$$

By substituting the value of rd in equation 8, then (TEC) is represented as follow:

$$TEC = (H\lambda)\times\left[\int k(v1,t)K(v) - k(v,t)K(v1))\sigma(v,v1)dv1\right]^{(9)} - \sum PI\log_2\frac{PI}{\sum_x\prod_{d=1}^{h-1}\binom{r_d + j_d - 1}{t_d}} - \sum_k - NI\log_2\frac{NI}{\sum_x\prod_{d=1}^{h-1}\binom{rd + j_d - 1}{t_d}}$$

On other words, the total exit complexity is represented in the following equation:

$$TEC = \frac{O^2}{2sa^2}\left[2U\lambda_n - \sum_{a=1}^{U}(\lambda_{n-a} + \lambda_{n+a})\right] \times \left[\int \begin{matrix} k(v1,t)K(v) \\ - k(v,t)K(v1))\sigma(v,v1)dv1 \end{matrix}\right]$$
$$-\sum PI \log_2 \frac{PI}{\sum_{x} \prod_{d=1}^{h-1} \binom{r_d + j_d - 1}{td}}$$
$$-\sum_{k} NI \log_2 \frac{NI}{\sum_{x} \prod_{d=1}^{h-1} \binom{r_d + j_d - 1}{t_d}}$$

Equation 9 represents the total exit complexity in quantum distributed systems considering all the operators that can affect this complexity. The algorithm depicted in table1 demonstrates the practical steps to calculate the complexity of knowledge in distributed quantum computer Systems. The algorithm depends on building bilateral non-tree entitled roots as a model of knowledge in distributed computing. In addition to some special accounts such as determining the length and the distance between each episode in the tree, and then switch between these episodes. We must put all storage nodes and the general structure of a binary tree. At this step has to be to search for positive and negative moments during the movement of knowledge. Then calculate the amount of information specific similarities in the system. In this case, the algorithm calculates the degree of complexity out information. It can then calculate the roots of the tree of others entitled, which was considered a step by calculating total egress complexity. The algorithm shows the penultimate step when calculating the degree of complexity. And forbids certain operations to compute the total exit complexity.

1- generate graph tree.
2- calculate the height of the tree.
3- moving between two adjacent nodes.
4- save the position of each node.
5- save the total schematic representation of the tree.
6- search for the total positive instances and the total negative instances.
7- measure the pseudo-information measure Is.
8- Calculate the information needed to egress from no exit node.
9- calculate the unlabeled rooted trees.
10 – measure the egress complexity.
11- measure the kinetic complexity.
12- Calculate the total exit complexity as
(kinetic complexity.+ egress complexity)

Table1 : the complexity algorithm of knowledge in distributed quantum computer systems

B. Minimizing the exit complexity in distributed quantum computing systems

In this section we discuss a methodology to minimize the exit complexity in distributed quantum computing systems. The main concept in minimizing the exit complexity is to portion the main system into subsystems logically. This portioning process contribute to reduce the different types of complexity in quantum systems. This concept of portioning is first introduced by Mingsheng Ying et al in [37]. Also they introduced basic algebraic laws for these groups of logical circuits using more detailed cases. This paper uses the algebraic formulas of division in subsystems and extend this work to study the effect of this division on the total exit complexity of the system. Mingsheng Ying et al in [37] consider CR as a circuit and $P = \{G_t : t \in E\}$ a partition of D(CR). They mention that C deference P when C deference that partition $P \cup \{G \setminus D(C)\}$. Where MQ1 and MQ2 be two sets of qubit names. If CR deference $P = \{G_t : t \in E\}$ and $G_i \subseteq M_{ti}$ (i=1,2). Then CR separates G1 from G2 both quantum measurements and unitary transformations can only be performed on local subsystems. Classical information extracted by a measurement on one subsystem can be passed to other subsystems. Also, entanglement resources are allowed to reside between different subsystems, and thus to connect them. In other words, many subsystems can share a single quantum resource in a distributed system. Now, consider TECn is the exit complexity for the n circuit, and Pn is the fraction of the exit complexity related to the main system for the n circuit. To calculate the complexity at n circuit , the formula of TECn is represented as in equation 10:

$$TEC_n = \frac{TEC}{P_n} (10)$$

Each circuit in the subsystem will have few numbers of eignvalues λ. From equation 9 it is concluded that:

$$\frac{\partial(TEC)}{\partial(\lambda)} = \left[\frac{2UO^2}{2sa^2 P_n} - \frac{\partial}{\partial\lambda}\sum_{a=1}^{U}(\lambda_{n-a} + \lambda_{n+a})\right] \times \left[\int \begin{matrix} k(v1,t)K(v) \\ - k(v,t)K(v1))\sigma(v,v1)dv1 \end{matrix}\right] = 0$$

Hence $$\frac{2UO^2}{2sa^2 P_n} = \frac{\partial}{\partial\lambda}\sum_{a=1}^{U}(\lambda_{n-a} + \lambda_{n+a})$$

$$P_n = \frac{2UO^2}{2sa^2\left(\frac{\partial}{\partial\lambda}\sum_{a=1}^{U}(\lambda_{n-a} + \lambda_{n+a})\right)} (11)$$

The formula of equation 11 presents the optimal value of Pn to obtain minimum values of the total exit complexity in logical subsystem. Table 2 depicts the algorithm used for An algorithm for minimizing the exit complexity in distributed quantum computing systems. It is depicted from this algorithm that is important steps to check various values of the circuits such as P , C and CR. This checking steps assure that the separation process is occurred.

1- Divide the main system logically into CR1,…..CRn

2- Check for C deference P

3- Check for P = { Gt : t,E}

4- Check that C deference P when C deference P union {G\D(CR).

5- If (CR deference P and Gi) Then CR separates MQ1 from MQ2)

6- Calculate TECn as (TEC/Pn)

7- Set the derivation of TEC with respect to λ to zero

8- Deduce the optimal value of Pn

1- Divide the main system logically into CR1,…..CRn
2- Check for C deference P
3- Check for P = { Gt : t,E}
4- Check that C deference P when C deference P union {G\D(CR).
5- If (CR deference P and Gi) Then CR separates MQ1 from MQ2)
6- Calculate TECn as (TEC/Pn)
7- Set the derivation of TEC with respect to λ to zero
8- Deduce the optimal value of Pn

Table2 : An algorithm for minimizing the exit complexity in distributed quantum computing systems

IV. System Simulation And Results

Figure 2 shows that when calculating the egress complexity at low values for both positive and negative moments of the rate increases tree logarithm of these moments in a linear fashion, regular and value very small egress complexity and increasingly up to scratch positive moments. In the mean values for positive moments increase egress complexity values, especially when the supreme moments as shown in Figure 3. When the supreme values of positive and negative moments value of egress complexity increases in the beginning, but when you increase the number of positive and negative moments significantly up to a value of zero as shown in Figure 4. Deduct from figures 2, 3 and 4 that the egress complexity be less than what can be at the highest values for both positive and negative moments. Figures 2, 3 and 4 refer to number of guidelines such as PI represented as s1, NI represented as s2 and log2(NI/m) represented as s3. Information is also evident from the figures 5, 6 and 7 that the value of the kinetic operator increases with increasing values of both h, m, a, a transaction affecting the overall value for kinetic complexity therefore increases significantly increased values of kinetic operator. figures 5, 6 and 7 refer number of guidelines such as h represented as s1, m represented s2, a represented as s3, N represented as s4, n represented as s5, NI represented as s6, PI represented as s7 and T represented as s8. Evidenced by the figures 8, 9 and 10 to exit complexity increases significantly when the supreme values of positive and negative moments of the tree that represents the volume of knowledge and significantly less at low values. Figures 8, 9 and 10 refer to the exit complexity as s1, egress complexity as s2 and kinetic operator as s3. For low values for exit complexity as shown in Figure 9 are working on the average values for higher values of positive and negative. From the above analysis this paper went to the importance of work to get to know on a regular basis with an average value for both the positive and negative moments to get the lowest value for the exit complexity. Practically These findings show the need to find a way to control the shape and structure of knowledge transmitted in distributed computing Quantity and control the speed of the transfer of this knowledge. Also end that we cannot separate the structure of knowledge and the complexity out of knowledge. The balanced structure of knowledge refers to the decline in values held out data from point to point in quantum systems.

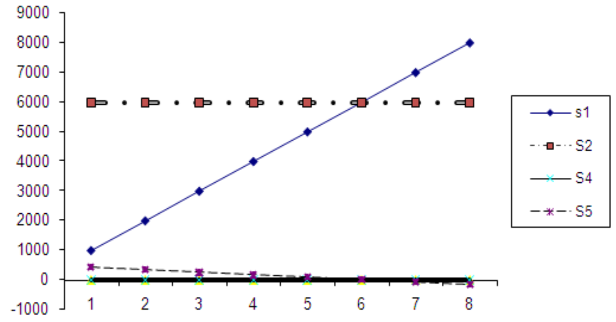


Fig.2 : Egress complexity at high values of PI and NI

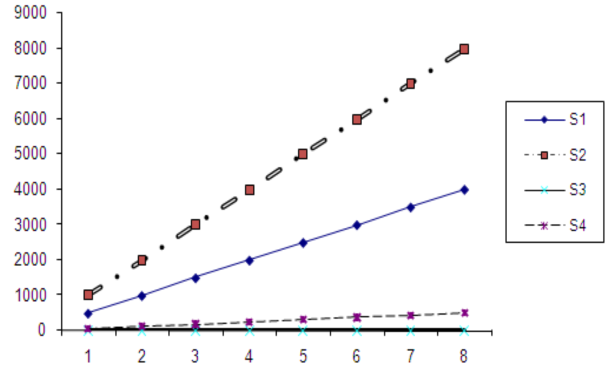


Fig.3 : Egress complexity at intermediate values of PI and NI

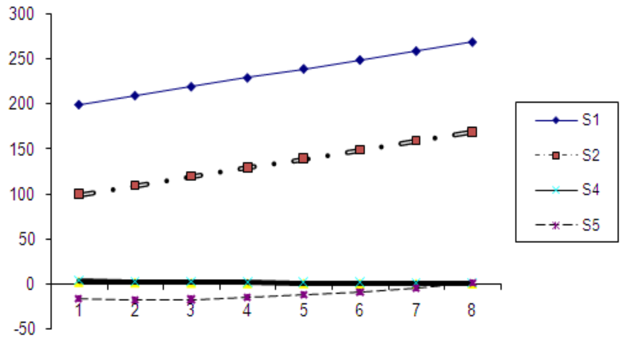


Fig.4 : Egress complexity at low values of PI and NI

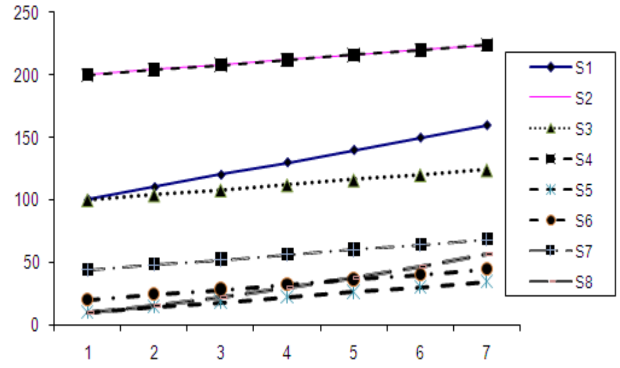


Fig.5 : kinetic operator at low values of NI and PI

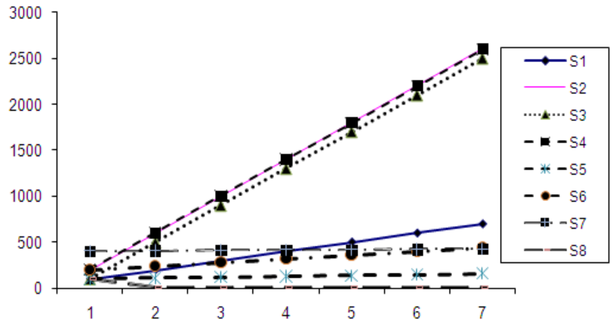


Fig. 6 : kinetic operator at intermediate values of NI and PI

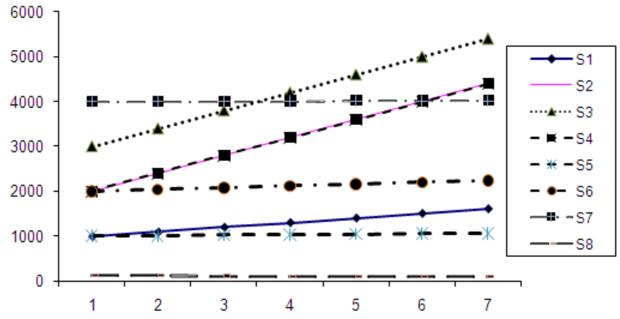


Fig.7 : kinetic operator at high values of NI and PI

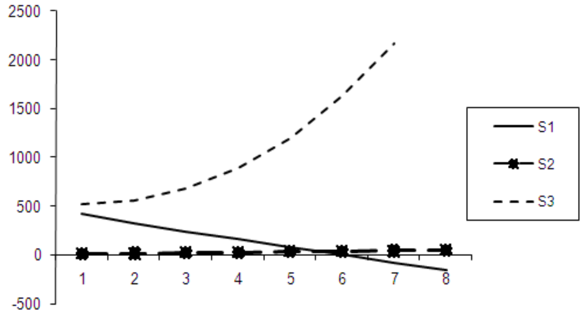


Fig.8 : Total Exit complexity at high values of egress complexity

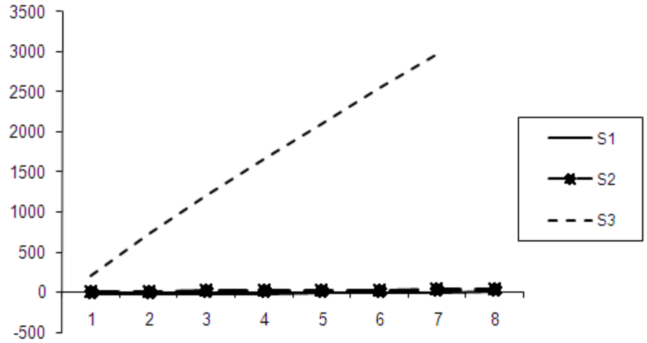


Fig.9 : Total Exit complexity at intermediate values of egress complexity

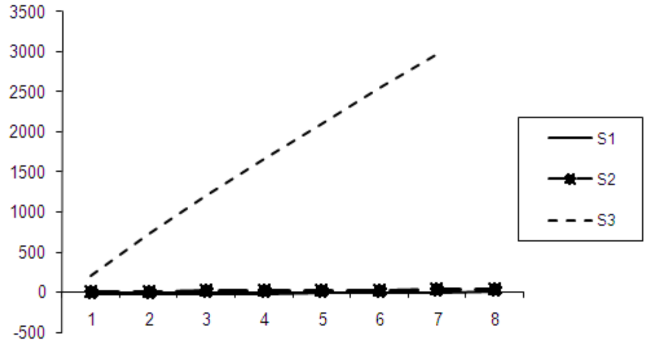


Fig.10 : Total Exit complexity at low values of egress complexity

V Conclusions

A quantum computer is a computer design which uses the principles of quantum physics to increase the computational power beyond what is attainable by a traditional computer. Quantum computers have been built on the small scale and work continues to upgrade them to more practical models. Quantum computer's main drawback is the same as its strength: quantum coherence. The qubit calculations are performed while the quantum wave function is in a state of superposition between states, which is what allows it to perform the calculations using both one and zero states simultaneously. All the above drawbacks can lead to more complexity in the exit of knowledge in distributed quantum computers. This research sheds light on how to calculate the degree of complexity out knowledge in distributed computing quantum systems. Then both are calculated Unlabeled binary trees In this research methodology proposed to express their knowledge by building. Total exit complexity and then calculated Kinetic complexity and Egress complexity. This paper also suggested using a logical approach to minimize the total exit complexity for the main system. The results indicate that the egress complexity be less than what can be at the highest values for both positive and negative moments. The transaction affecting the overall value for kinetic complexity therefore increases significantly increased values of kinetic operator. The exit complexity increases significantly when the supreme values of positive and negative moments of the tree that represents the volume of knowledge and significantly less at low values.

References

[1] http://en.wikipedia.org/wiki/Quantum_computer , visited 14-7-2012

[2] Ellie D’Hondt , " Distributed quantum computation a measurement-based approach", Faculteit Wetenschappen , July 2005

[3] Samson Abramsky & Bob Coecke. A categorical semantics of quantum protocols. In Proceedings of the 19th annual IEEE Symposium on Logic (LICS) in Computer Science. IEEE Computer Society, 2004.

[4] Pedro Ad˜ao & Paulo Mateus. A process algebra for reasoning about quantum security. In Peter Selinger, Ed., Proceedings of the 3rd Workshop on Quantum Programming Languages (QPL04), pages 3–20, 2005.

[5] Panos Aliferis & Debbie W. Leung. Computation by measurements: a unifying picture. Phys. Rev. A, 70:062314, 2004.

[6] Thorsten Altenkirch & Jonathan Grattage. A functional quantum programming language. In Proceedings of the 20th Annual IEEE Symposium on Logic in Computer Science (LICS). IEEE Computer Society, 2005.

[7] Andris Ambainis, Harry Buhrman, Yevgeni Dodis, & Hein R¨ohrig. Multiparty quantum coin flipping. In Proceedings of the Conference on Computational Complexity (CCC), 2004.

[8] Dana Angluin. Local and global properties in networks of processors. In Proceedings of the 12th annual ACM symposium on Theory of computing, pages 82–93. ACM Press, 1980.

[9] Alexandru Baltag & Sonja Smets. Quantum dynamic logic. In Peter Selinger, Ed., Proceedings of the 2nd Workshop on Quantum Programming Languages (QPL04), Turku, Finland, 2004. Turku Centre for Computer Science, TUCS General Publication No 33.

[10] Alexandru Baltag & Sonja Smets. LQP: the dynamic logic of quantum information. Unpublished, 2005.

[11] Simon C. Benjamin, Daniel E. Browne, Joe Fitzsimons, & John J.L. Morton. Brokered graph state quantum computing. 2005.

[12] Abrams D S and Lloyd S 1997 Simulation of many-body Fermi systems on a universal quantum computer Phys. Rev. Lett. 79 2586–9

[13] Aharonov D and Ben-Or M 1996 Fault-tolerant quantum computation with constant error Preprint quantph/ 9611025

[14] Aspect A, Dalibard J and Roger G 1982 Experimental test of Bell’s inequalities using time-varying analysers Phys. Rev. Lett. 49 1804–7

[15] Barenco A 1995 A universal two-bit gate for quantum computation Proc. R. Soc. A 449 679–83

[16] Barenco A, Bennett C H, Cleve R, DiVincenzo D P, Margolus N, Shor P, Sleator T, Smolin J A and Weinfurter H 1995b Elementary gates for quantum computation Phys. Rev. A 52 3457–67

[17] Speakable and Unspeakable in Quantum Mechanics (Cambridge: Cambridge University Press) Benioff P 1980 J. Stat. Phys. 22 563

[18] Bennett C H, Bernstein E, Brassard G and Vazirani U 1997 Strengths and weaknesses of quantum computing Preprint quant-ph/9701001

[19] Bennett C H, Brassard G, Briedbart S and Wiesner S 1982 Quantum cryptography, or unforgeable subway tokens Advances in Cryptology: Proceedings of Crypto ’82 (New York: Plenum) pp 267–75

[20] Berman G P, Doolen G D, Holm D D, Tsifrinovich V I 1994 Quantum computer on a class of one-dimensional Ising systems Phys. Lett. 193 444–50

[21] Bernstein E and Vazirani U 1993 Quantum complexity theory Proc. 25th Annual ACM Symposium on Theory of Computing (New York: ACM) pp 11–20

[22] Berthiaume A and Brassard G 1992a The quantum challenge to structural complexity theory Proc. 7th Annual Structure in Complexity Theory Conf. (Los Alamitos, CA: IEEE Computer Society Press) pp 132–7

[23] Braunstein S L, Mann A and Revzen M 1992 Maximal violation of Bell inequalities for mixed states Phys. Rev. Lett. 68 3259–61

[24] Brune M, Nussenzveig P, Schmidt-Kaler F, Bernardot F, Maali A, Raimond J M and Haroche S 1994 From Lamb shift to light shifts: vacuum and subphoton cavity fields measured by atomic phase sensitive detection Phys.

[25] qubit-external.physics.ox.ac.uk/...quantum-computing/.../20-quantum-computing-longer-intro.htm , visited 24-8-2012

[26] Simon D , "On the power of quantum computation", in Proc. 35th Annual Symposium on Foundations of Computer Science (IEEE Computer Society Press, Los Alamitos), 1994 , 124-134

[27] Shor P W ," Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer", in Proc. 35th Annual Symp. on Foundations of Computer Science, Santa Fe, IEEE Computer Society Press; revised version 1995a preprint quantph/ 9508027, 1994

[28] Kitaev A Yu," Quantum measurements and the Abelian stablizer problem", (preprint quantph/ 9511026), 1995

[29] Beckman D, Chari A, Devabhaktuni S and Preskill J " Efficient networks for quantum factoring", Phys. Rev. A 54, 1034-1063, 1996

[30] Alexandru Baltag & Sonja Smets. "Quantum dynamic logic". In Peter Selinger, Ed., Proceedings of the 2nd Workshop on Quantum Programming Languages (QPL04), Turku, Finland, 2004. Turku Centre for Computer Science, TUCS General Publication No 33.

[31] Vincent Danos, Ellie D’Hondt, Elham Kashefi, & Prakash Panangaden." Distributed measurement-based quantum computation". In Peter Selinger, Ed., Proceedings of the 3rd Workshop on Quantum Programming Languages (QPL05), 2005. quant-ph/0506070.

[32] Ron van der Meyden & Manas Patra. Knowledge

in quantum systems. In Proceedings of the 9th conference on Theoretical aspects of rationality and knowledge, pages 104–117, Bloomington, Indiana, June 2003.

[33] Hongjun Jiang, "The development of a scenario Independent Method for Evaluating the Evacuation Complexity of a building", March 2012 , The university of Greenwish

[34] H. A. DONEGAN, "Formal Aspects of Egress Complexity", Mathematical and Computer Modelling 35 (2002) 119-128

[35] Gang Li , " Generation of Rooted Trees and Free Trees", University of Victoria 1996.

[36] Christian Lécot et al , " Quasi-random Simulation of Linear Kinetic Equations" , journal of complexity 17, 795–814 (2001)

[37] Mingsheng Ying and Yuan Feng, "An Algebraic Language for Distributed Quantum Computing" IEEE Transactions on Computers, VOL. 58, NO. 6, JUNE 2009

Study of Proper Hierarchical Graphs on a Grid

Mohamed A. El-Sayed
Dept. of Mathematics, Faculty of Science,
University of Fayoum, Egypt.
Assistant professor of CS, Taif University, KSA

Ahmed A. A. Radwan,
Dept. of Computer Science, Faculty of Science,
Minia University, Minia, Egypt

Nahla F. Omran
Dept. of Mathematics, Faculty of Science,
South Valley University, Qena, Egypt

***Abstract*—Hierarchical planar graph embedding (sometimes called level planar graphs) is widely recognized as a very important task in diverse fields of research and development. Given a proper hierarchical planar graph, we want to find a geometric position of every vertex (layout) in a straight-line grid drawing without any edge-intersection. An additional objective is to minimize the area of the rectangular grid in which *G* is drawn with more aesthetic embedding. In this paper we propose several ideas to find an embedding of *G* in a rectangular grid with area, (ℓ-1) × (*k*-1), where ℓ is the number of vertices in the longest level and *k* is the number of levels in *G*.)**

Keywords-level graphs; hierarchical graphs; algorithms; graph drawing.

I. INTRODUCTION

The drawing of directed acyclic graphs (DAG) is widely recognized as a very important task in diverse fields of research and development. Examples include VLSI Design and plant layout [1], graphical user interfaces [2], software and information engineering, project management, visual languages [3], subroutine-call graphs, Interpretative Structural Modeling [4], organization charts, hierarchical relationships, system theory and other research fields. The usefulness of a graph depends on its layout that should be readable, understandable and easy to remember. A fundamental issue in Automatic Graph Drawing is to display hierarchical network structures, as they appear in many applications. The network is transformed into a directed acyclic graph (DAG) that has to be drawn with edges that are strictly monotone with respect to the vertical direction. Many applications imply a partition of the vertices into levels that have to be visualized by placing the vertices that belonging to the same level on a horizontal line. The corresponding graphs are called level graphs, and the drawing of the networks that correspond to this category of graphs means the drawing of level graphs (see [5]).

The use of integer coordinates in embedding a graph on the grid has many advantages such as speed, accuracy, and it guarantees automatically that the resultant picture has fairly good properties. A straight line drawing is a grid drawing if each vertex is at a grid point, and the edges are represented as straight-line segments between their endpoints without any edge-intersection. See Figure 1.

Usually, we use one of some aesthetic criteria (such as drawing area minimization, minimizing the number of edge crossings, symmetry, bends minimization or distributing the vertices uniformly) in order to make the layout of a graph readable and understandable [6],[7]. Reducing the number of edge crossings or distributing the vertices uniformly have been proposed and evaluating goodness of drawing based on these criteria has been reported [8], [9], [10]. Many works have been published area requirements for drawing hierarchically planar graphs [6],[7], [8], [11].

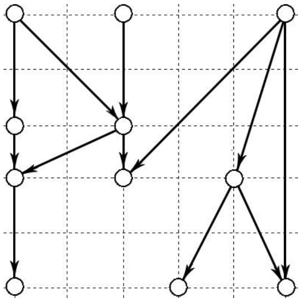


Figure 1. Straight line drawing of level planar graph

In [6] introduced genetic algorithms (GAs) with the problem of drawing of level planar graph or hierarchical planar graph, and explored the potential use of GAs to solve this particular problem. They showed that the GAs can help find a layout of levels and hierarchical planar graphs without any crossing edges.

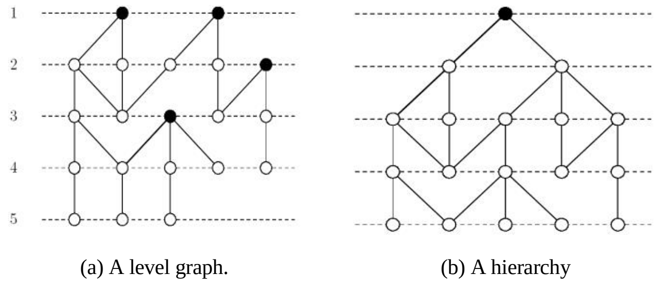
Lin and P. Eades [12] show that for each hierarchically planar straight-line drawing of G, where each pair of vertices in the same layer are at least distance 1 apart, has width at least $\Omega((2n-1)!)$ where n is the number of vertices in the graph. J. Abello [13] highlighted the main tasks behind the computation of hierarchical graph maps and provided several examples. The techniques have been used experimentally in the navigation of very large graphs.

In this paper, we are concerned with drawing of level planar *G* in the plane such that the vertices of *G* are represented as grid points, and the edges are represented as straight-line segments between their endpoints without any edge-intersection. An additional objective is to minimize the area of the rectangular grid in which *G* is drawn. We introduce new algorithms these find an embedding of *G*. These new algorithms gives level planar drawing in a rectangular grid with area (ℓ -1)×(*k*-1), where ℓ is the number of vertices in the longest level.

This paper is organized as follows. After summarizing the necessary preliminaries in the next section. In the first algorithm, *SimpleProperLevel*, we explain proper placement for the vertices of a level graph with minimum area in the third section. In the fourth section we explain the second algorithm, *FixedDistance*, redistribute of the vertices with fixed distance between any two consecutive vertices in the same level on the grid drawing. In section five, *DegreeDistance* algorithm, redistribute of the vertices on the grid according to its degrees. Finally, Conclusion and references are presented.

II. Preliminaries

Given a directed acyclic graph $G = (V, E, \phi)$. A *leveling* of *G* is a topological numbering of *G*, where $\phi : V \rightarrow Z$ mapping the vertices of *G* to integers such that $\phi(v) \geq \phi(u) + 1$ for all $(u, v) \in E$. *G* is called a *level graph* if it has a leveling. If $\phi(v) = j$, then *v* is a level-*j* vertex. Let $V^j = \phi^{-1}(j)$ denote the set of level-*j* vertices. Each V^j is a level of *G*. If $G = (V, E, \phi)$ has a leveling with *k* being the largest integer such that V^k is not empty, *G* is said to be a *k*-level graph. For a *k*-level graph *G*, we sometimes write $G = (V^1, V^2, ..., V^k; E)$.

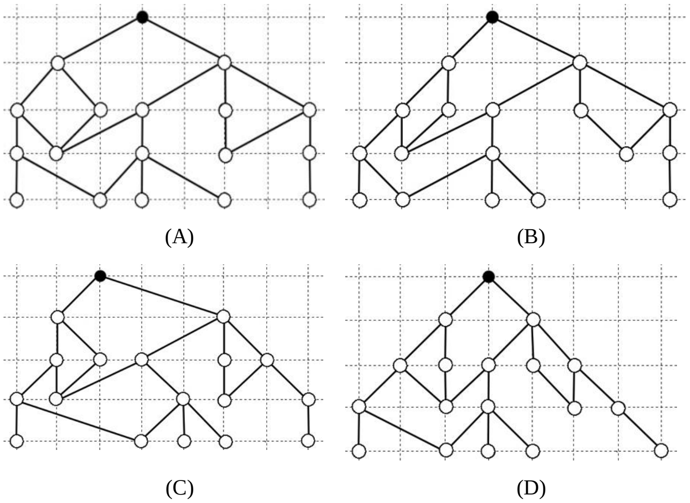


(a) A level graph.

(b) A hierarchy

Figure 2. Examples of proper level graphs, sources are drawn black

A drawing of G in the plane is a level drawing if the vertices of every V^j, 1 ≤ j≤ k, are placed on a horizontal line $l_j = \{(x, k-j) | x \in R\}$, and every edge $(u, v) \in E$, $u \in V^i$, $v \in V^j$, $1 \leq i < j \leq k$, is drawn as a monotone decreasing curve between the lines l_i and l_j. A level drawing of G is called level planar if no two edges cross except at common endpoints. A level graph is level planar if it has a level planar drawing. A level graph $G = (V, E, \phi)$ is said to be proper, if every edge $e \in E$ connects only vertices belonging to consecutive levels. Usually, level graph G has sinks and sources placed on various levels of the graph. For example, Figure 2.a, taken from [14],[15], shows a level graph. Hierarchy is a level graph such that all sources belong to the first level V^1 of the graph. As a consequence, we consider only hierarchies with $|V^1| = 1$. Figure 2.b shows a hierarchy.



(A)

(B)

(C)

(D)

Figure 3. Four different embeddings of the same level graph in area 7x4

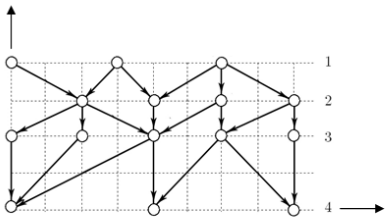
In Figure 3 we give four different embeddings of the same level graph in a rectangular grid with area 7×4. From Figure 3, it is so easy to observe that for a level graph, there are several embeddings; each one differs in view from the others. The difference in view between several embeddings of the same graph is due to the differences in the distances between each two consecutive levels in the graph, and to the differences in the distances between each two consecutive vertices in each level. According to a certain application, an embedding of a level graph may be more convenient than the other ones, and the convenient embedding may be inconvenient to another application.

Jünger, Liepert and Mutzel [16] have given a level planarity test algorithm of *G* in linear time. Using PQ-tree data structure, Jünger and Liepert [5] have given an algorithm that embeds a level planar graph in linear time, that algorithm was based on a level planarity test in [16].

By *P*(*v*) we will denote the current position of vertex *v* in the grid, i.e., *P*(*v*):=(*x*(*v*),*y*(*v*)). By *P*(*u*,*v*) we denote the embedding of edge *e*(*u*,*v*), that is, the line segment that connects *P*(*u*) with *P*(*v*). The following symbols will be used in this paper:

h_I	is the vertical distance between any two consecutive levels in the graph.
d_i	is the horizontal distance between any two consecutive vertices in the level V^i.
D	is the horizontal distance between any two consecutive vertices in the longest level.
l_I	is the number of vertices that belong to the level V^i.
ℓ	is the number of vertices in the longest level.
K	is the number of levels in given level graph *G*.
Δ_i	is the total degree of the vertices in the level
δ_j	is the degree of the vertex j in the level
H,*W*	are the height and the width of the used rectangular grid, respectively.

In Figure 4 an example to illustrate some of above symbols of given proper level graph, with four levels. The longest level is V^3 with five vertices.



i	h_i	d_i	l_i	Δ_i
1	1	3	3	6
2	1	2	4	14
3	2	2=d	5=ℓ	15
4	-	4	3	7
H=4, W=8				

Figure 4. Example of proper level graph k=4.

III. Embedding of Graph in Minimum Area

Let $G = (V^1, V^2, ..., V^k; E)$ be a given proper level graph with n vertices. An embedding of a level graph can be aesthetic if we redraw the graph by making the distances between each two consecutive vertices, d_i are equal in the same level, and for every level the distances between each two consecutive levels, h_i are equal also.

Now we describe the embedding strategy of the first algorithm. Simply, we put d_i =1 and h_i =1 for all levels, but the positions of vertices for each level distributed about the middle point of the width W at this level. The value [W / 2] means that, the integer part of real value W/2. It is clear that, the high of the used rectangular grid, $H = h_1 + h_2 + \dots h_{k-1} = k-1$ and the width is the total distances d_i in the longest level. i.e $W = \Sigma\, d_i = \ell - 1$.

The complete *SimpleProperLevel* ($G(V, E, \phi)$) algorithm can now be described as follows:

```
Algorithm SimpleProperLevel ( $G(V, E, \phi)$ );
  Input: A given level planar graph $G(V^1, V^2, ..., V^k; E)$ with $n$ vertices.
  Output: An embedding of the level graph $G$ on grid drawing.
   Begin
     Let $\ell$ is the number of vertices in the longest level.
     Now we compute the $x$- and $y$-coordinate of
     $V^i = (v_1, v_2, ..., v_{l_i}), i = 1,2,...,k$ , as follows:
        For $i$=1 to $k$
        Begin
           Let $V^i = (v_1, v_2, ..., v_{l_i})$ . $l_i$ is the number of
           vertices in level $i$ .
                 For $j$=1 to $l_i$
                 Begin
                       $y(v_j) = k - i$                              (1)
                       $x(v_j) = [\ell / 2] - [l_i / 2] + j - 1$     (2)
                 End
                 { Now we have a drawing of $G_i$.}
        End
        Output the drawing of $G$.
```

End.

Lemma 1: *For each* 1< $i \le k$, *when we add* V^i, *then after applying the equations (1)&(2), all neighbors of* V^i *are visible, that the edges between* V^i *and* V^{i-1} *do not intersect themselves.*

Proof: Since $G = (V^1, V^2, ..., V^k; E)$ be a given proper level graph. All neighbors of $v \in V^i$ in G_{i-1} are only in the previous level V^{i-1}. The y-coordinate value is k- i +1 of $v' \in V^{i-1}$ and k-i of any vertex in the consecutive level V^i. So, all neighbors of $v \in V^i$ are visible. Since the vertices in the levels are ordered, The x-coordinate values of $V^i = (v_1, v_2, ..., v_{l_i})$ are determined by equation (2) according to its order in V^i, so the edges between V^i and V^{i-1} do not intersect.

The lemma above implies immediately that adding V^i,1< $i \le k$, satisfies the conditions of drawing level graph, as stated in the corollary below.

Corollary 1: *For each* 1< $i \le k$, *The sub-graphs* $G' = (V^1, V^2, ..., V^i; E')$ *of G remain proper level graph during the algorithm*.

Theorem 1: *SimpleProperLevel algorithm constructs a straight-line embedding of any proper level graph* $G(V^1, V^2, ..., V^k; E)$ *into a* (ℓ -1)×(k-1) *grid, where* ℓ *is the number of vertices in the longest level in G.*

Proof: It is clear that, the height H of drawing grid, $H = h_1 + h_2 + \dots + h_{k-1}$, since h_i =1 as minimum value for all levels, then $H = k-1$. Also, from equation (1), $H = y(v) = k-1$, for any $v \in V^1$. The width is equal to x-coordinate value of vertex number ℓ in the longest level. Replace l_i by ℓ in equation (2), $W = x(v_\ell) = [\ell/2] - [\ell/2] + \ell - 1 = \ell - 1$. Hence, the area used for drawing any given proper level graph is a function of the number of levels and the number of vertices in the longest level. Hence, the proof is completed.

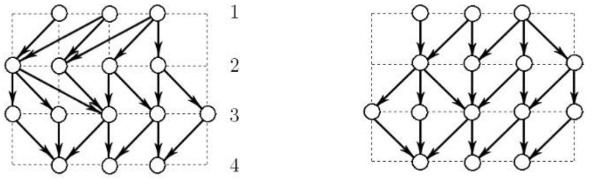
Theorem 2: *Let* $G(V^1, V^2, ..., V^k; E)$ *be given, then SimpleProperLevel algorithm can be computed in linear-time.*

Proof: We embed one level every one-time run of outer-loop in the algorithm. Since k is the number of levels in given level graph G, outer-loop is run in k-times. In inner-loop, every vertex in a level will be visited once. But l_i is number of vertices in a level V^i. Hence the algorithm can be computed in $\sum_{i=1}^{k} \sum_{j=1}^{l_i} v_j = n$ times. Hence, the proof is completed.

Notice: We can get more aesthetic embedding of any level graph G by replacing the equation (2) by the following equation:

$$x(v_j) = [(\ell + 1)/2] - [(l_i + 1)/2] + j - 1 \qquad (3)$$

We can see the difference between using *SimpleProperLevel* algorithm with the two equations (1&2) and with the two equations (1&3), by embedding the given graph in figure 4. It is clear that the embedding of Figure (5.b) is more aesthetic than the one given in Figure (5.a).



(a) using equations (1&2). (b) using equations (1&3)

Figure 5. Example of proper level grap using *SimpleProperLevel algorithm*

IV. Distributing the Vertices with Constant Distances

Let $G=(V^1,V^2,...,V^k;E)$ be a given level planar graph with *n* vertices. The values of the distance d_i between each two consecutive vertices in a level are computed by dividing the number of vertices in the longest level ℓ by the number of vertices in that level *i*. Its means that d_i= [ℓ / l_i]. Since ℓ is greater than or equal to l_i , then we ensure that d_i is at least one grid unit . Note that the distance d_i between each two consecutive vertices is constant value through level *i*. In order to determine $P(v_j)$, j=1,2,…,l_i , in the proposed algorithm , when adding a vertex v_j , we determine its location in the grid by placing v_j such that $P(v_j)=(x, y)$, where:

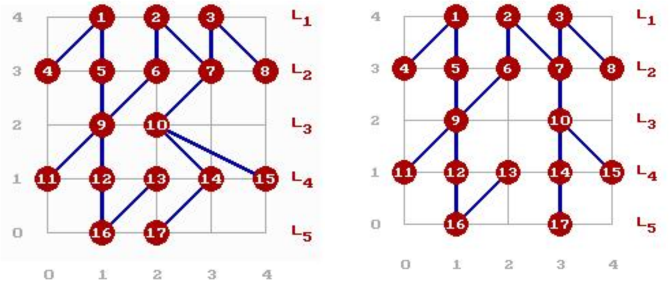
$$y(v_j)= k- i$$

$$x(v_j)= [\ell / 2 - (l_i -1)*d_i / 2] + (j - 1)*d_i \qquad (4)$$

In the equation (4), we will locate the leftmost and rightmost vertices in the longest level to the left and right boundaries of the grid respectively. And for the other levels we keep to equal distances, one between the left boundary and the leftmost vertex and the other one between the right boundary and the rightmost vertex using the term [(ℓ / 2 - (l_i –1)*d_i / 2]. The output of this algorithm is an embedding of the level graph *G* on grid drawing with constant distance d_i and more visible distribution of vertices than the *SimpleProperLevel* algorithm. The complete *ConstantDistance*($G(V, E,\phi)$) algorithm can now be described as follows:

```
Algorithm ConstantDistance ( $G(V, E,\phi)$ );
  Input: A given $G(V^1,V^2,...,V^k;E)$ with n vertices.
  Output: An embedding of G on grid drawing.
    Begin
      For i=1 to k
      Begin
            $d_i$= [ $\ell$ / $l_i$ ]
            For j=1 to $l_i$
            Begin
               $y(v_j)= k- i$
               $x(v_j)= [ \ell / 2 - (l_i -1)*d_i / 2] + ( j - 1)*d_i$
            End
      End
    Output the drawing of G.
  End.
```

Here we give an example to compare the above algorithms. In this example, we embed a given level graph with seventeen vertex (*n=17*), sixteen edge (*m=16*), five levels (*k=5*), and 5 vertices in the longest level which is the level number 2 or 4, (ℓ *=5*). We embed it using *SimpleProperLevel* algorithm, where the distance between any two vertices is one unit, see Figure (6.a). Figure (6.b) shows a new embedding of the same level graph after applying of *ConstantDistance* algorithm, where according to number of vertices in a level the distance d_i is computed.



(a) *SimpleProperLevel* algorithm. (b) *ConstantDistance* algorithm

Figure 6. Illustration of the distribution of vertices on grid

Theorem 3: *ConstantDistance algorithm constructs a straight-line embedding of level graph* $G(V^1,V^2,...,V^k;E)$ *into a* (ℓ -1)×(*k*-1) *grid.*

Proof: It is clear that, the high of drawing grid, $H = y(v \in V^1)$ = *k*-1. The width is equal to *x*-coordinate value of vertex ℓ in the longest level. Since d_i= [ℓ / l_i], put $l_i = \ell$ in equation (4), then we obtain that $W = x(v_\ell)$= [ℓ / 2 - (ℓ –1)*d_i / 2] + (*j* – 1)*d_i = [ℓ / 2 - (ℓ –1)*1 / 2] + (ℓ – 1)*1 = ℓ – 1. Hence, the proof is completed.

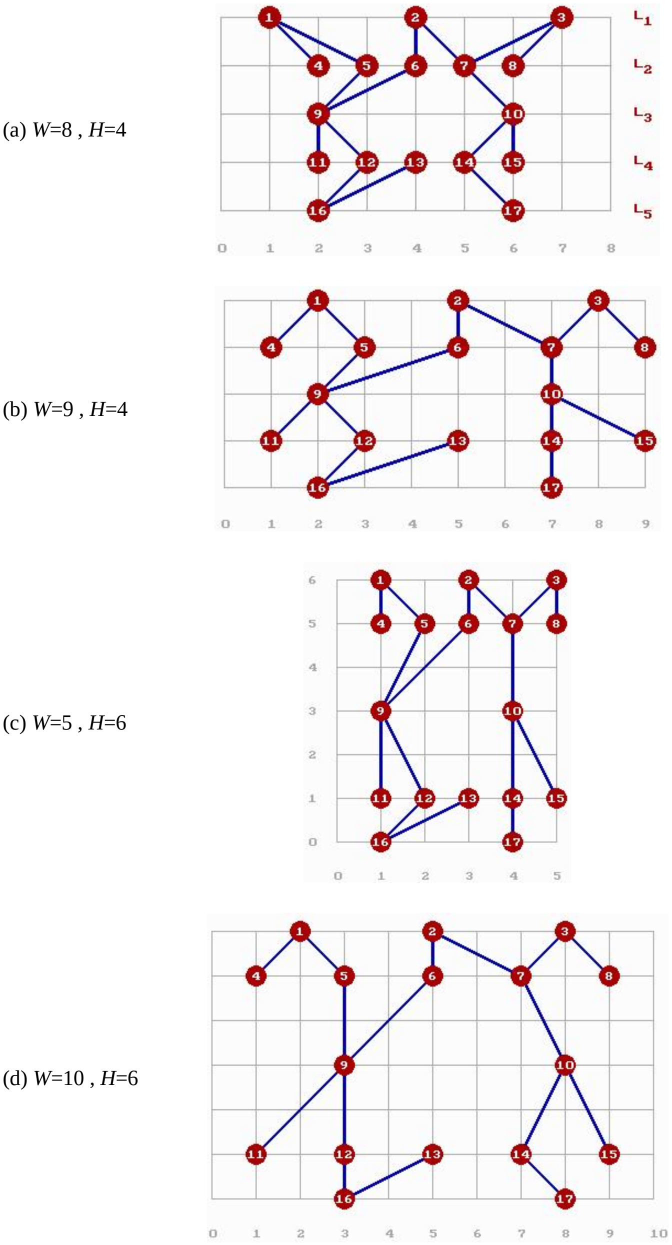
Theorem 4: *Let* $G(V^1,V^2,...,V^k;E)$ *be a given , then ConstantDistance algorithm can be computed in linear-time.*

Lemma 2: *The proper level graph* $G(V^1,V^2,...,V^k;E)$ *can be embedded in any constant area, such that* $W \geq \ell -1$ *and* $H \geq k-1$.

Proof: Using *ConstantDistance* algorithm, Scince $W= \ell -1$ as minimum width in grid, then we can take d_i= [(*W*+1)/l_i] for $W \geq \ell -1$ and the equation (4) become

$$x(v_j)= [(W+1)/ 2 - (l_i -1)*d_i / 2] + (j - 1)*d_i$$

Applying the equation (4), we ensure that all vertices are embeds in the width *W*. On the other hand, the hight is arbitrary positive integer value between any two levels in level graph *G*. The proof is completed.



(a) W=8 , H=4

(b) W=9 , H=4

(c) W=5 , H=6

(d) W=10 , H=6

Figure 7. Embedding of the same level graph in Figure 6 in 4 different areas.

Figure 7 illustrate lemma 2, embedding of the same level graph in Figure 6 in four different areas. It is so easy to observe that for a level graph, there are several embeddings, each one differs in view from the others and the convenient embedding may be inconvenient to another application.

V. Distribution on Grid Using Degrees of Vertices

In this section we distribute the vertices on the grid according to its degree. So, the distance between each two consecutive vertices is not fixed value in the same level i. Let Δ_i is the total degrees of the vertices in the level i, and δ_j is the degree of the vertex j. We calculate the weight ξ_j of each vertex in any level i, as the average value of the vertex v_j degree and the degress of its left and right vertices, $\xi_j = (\delta_{j-1} + \delta_j + \delta_{j+1})/3$. Note that if there is not any left or right vertex we consider that the degree of left or right equal to zero. Hence, we can calculate the distance d_{ij} between the vertex $v_{j\text{-}1}$ and v_j as a function of total degree Δ_i of the level and its weight ξ_j, so we can put d_{ij}= [$\ell * \xi_j / \Delta_i$]. To overcome confidingness, if $d_{ij} = 0$, $j > 1$, we consider d_{ij} = 1. In this case, when adding a vertex v_j, we determine its location in the grid by placing v_j such that $P(v_j)=(x, y)$, where:

$$y(v_j)= k\text{- } i$$

$$x(v_j)= x(v_{j\text{-}1})+ d_{ij}\,,\ \text{where } d_{ij}= [\,\ell * \xi_j / \Delta_i\,] \qquad (5)$$

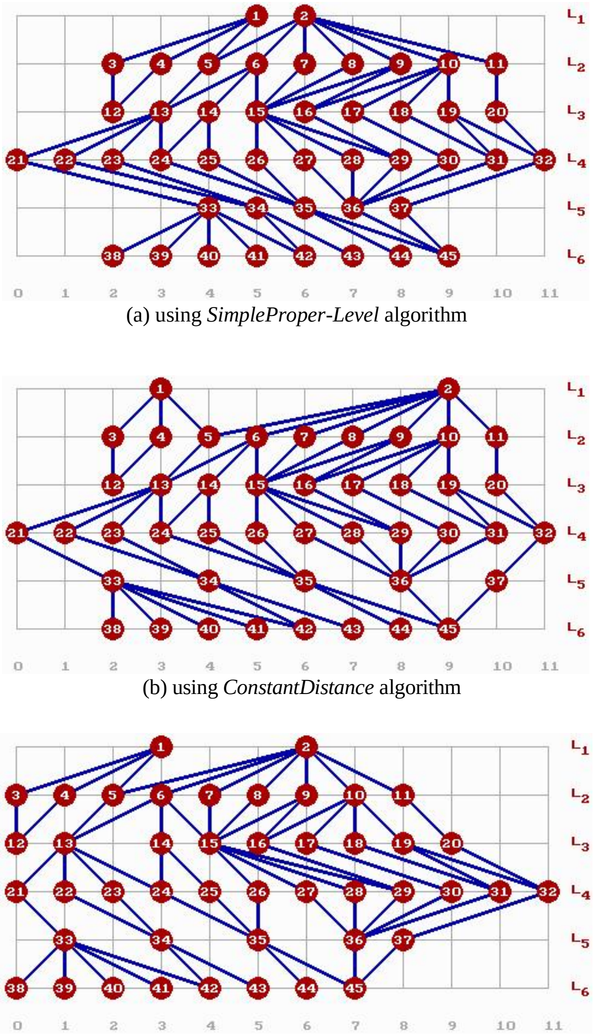
The complete *DegreeDistance*($G(V, E, \phi)$) algorithm can now be described as follows:

```
Algorithm DegreeDistance($G(V, E, \phi)$);
 Input: A given $G(V^1, V^2, ..., V^k; E)$ with $n$ vertices.
 Output: An embedding of the level graph G on grid drawing.
  Begin
     For $i$=1 to $k$
     Begin
          For $j$=1 to $l_i$
          Begin
               $d_{ij}$= [ $\ell * \xi_j / \Delta_i$ ];
               If ($d_{ij}$= 0) and ( $j$>1) Then $d_{ij}$= 1;
               $x(v_j)= x(v_{j-1}) + d_{ij}$ ;
               $y(v_j)= k- i$
          End
      End
      Output the drawing of G.
   End.
```

Consequently, from above algorithm, we obtain the following theorem:

Theorem 5: *DegreeDistance algorithm constructs a straight-line embedding of proper level graph* $G(V^1, V^2, ..., V^k; E)$, *which needs a rectangular grid with area at least* (ℓ -1) × (k-1) *grid and it can be computed in linear-time.*

Proof: It is clear that, the high of drawing grid, $H = y(v \in V^1) = k\text{-}1$. At least the width is equal to x-coordinate value of vertex ℓ in the longest level. Since $d_{ij} \geq 1$, From equation (5), we obtain that $W = \Sigma\, d_{\ell j} \geq \ell - 1$. Then in this case, embedding of proper level graph $G(V^1, V^2, ..., V^k; E)$, which needs a rectangular grid with area at least (ℓ -1)×(k-1) grid. Since we embed one level every one-time run of outer-loop in the algorithm. And k is the number of levels in given level graph G, outer-loop is run in k-times. In inner-loop, every vertex in a level will be visited once in a level V^i. Hence the algorithm can be computed in linear-time times. Hence, the proof is completed.



(a) using *SimpleProper-Level* algorithm

(b) using *ConstantDistance* algorithm

(c) using *DegreeDistance* algorithm

Figure 8. Three different embeddings of the same graph in minimum area.

Here we give an example to compare the above three algorithms. In this example, we embed a given level graph with n=45, m=64, six levels (k=6), and twelve vertices in the longest level which is the level number 4, (ℓ =12). In Figure 8.(a) using *SimpleProper-Level* algorithm. The output of this drawing has the property that the distance between any two vertices for all vertices is constant and equal to one grid unit. Figure 8.(b) using *ConstantDistance* algorithm. This drawing has the property that the distance between any two vertices in the same level are constant and is equal to one or more one grid unit. In Figure 8.(c) using *DegreeDistance* algorithm. The output of this drawing depends on the degree of the vertex and its neighbors for all vertices.

VI. Conclusion

In this paper, we introduced new three algorithms for embedding a proper level graph on a grid with minimum width. These algorithms keep a proper placement for the nodes that belonging to the same level for each level in the graph. An additional objective is to minimize the area of the rectangular grid in which G is drawn with more aesthetic embedding, that is clearly in the second algorithm. These algorithms run in time $O(n)$ where n is the number of nodes in the graph. It is possible to draw a given level graph within the area W×H. One of the goals of this area of research should be to extend this further, and to determine an optimal width-height tradeoff for grid drawings. Then for any feasible pair (W, H) we could apply a method that gives best drawings on grids of size (W, H). In this paper we introduced several ideas to find an embedding of G in a rectangular grid with area, (ℓ-1)×(k-1), where ℓ is the number of vertices in the longest level and k is the number of levels in G.

Acknowledgment

The authors wish to thank Prof. Abdel Badeeh Salem for his valuable hints and comments.

References

[1] F. Chung, F. Leighton and A. Rosenberg, “Embedding Graphs in Books: A Layout Problem with Applications to VLSI Design”, SIAM Journal Discrete Mathematics, 8, pp. 33-58, 1987.

[2] G. Di Battista, P. Eades, R. Tamassia and I. Tollis, “Algorithms for Drawing Graphs: Annotated Bibliography”, Computational Geometry, Theory Applications, 4, pp. 235-282, 1994.

[3] S. Bhatt and F. Leighton, “A Framework for Solving VLSI Graph Layout Problems”, Journal of Computer and System Systems Sciences 28, pp. 300-343, 1984

[4] P. Eades and X. Lin, “How to Draw a Directed Graph”, Proceeding IEEE on Visual Languages, (VL ’89), pp. 13-17, 1989.

[5] M. Jünger and S. Liepert, "Level Planar Embedding in Linear Time", Technical Report 99-374, Institut für Informatik, Unversität zu Köln, 1999.

[6] Ahmed A. A. Radwan, Mohamed A. El-Sayed and Nahla F. Omran, “Hybrid GA for Straight-Line Drawings of Level Clustered Planar Graphs”, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 2, pp. 229- 235, July 2011.

[7] A. Yamaguchi and A. Sugimoto, “An Approximation Algorithm for the Two-Layered Graph Drawing Problems”, COCOON ’99, (Lecture Notes in Computer Sciences, 1627), pp. 81-91, 1991.

[8] C. Batini, L. Furlani and E. Nardelli, “What is a Good Diagram? A Pragmatic Approach”, Proceeding Of the 4th International Conference on Entity-Relationship Approach, pp. 312-319, 1985.

[9] C. Esposito, “Graph Graphics: Theory and Practice”, Computer and Mathematics with Applications 15 (4), pp. 247-253, 1988.

[10] H. Purchase, R. Cohen and M. James, “Validating Graph Drawing Aesthetics”, Proceeding of Symposium on Graph Drawing, GD ’95 (Lecture Notes in Computer Sciences, 1027), pp. 435-446, Springer, 1996.

[11] A. A. A. Radwan, M. R. Girgis and A. A. Ghanem, “A Study of the Area of Drawing Level Graphs”, International Journal of Applied Mathematics, 5(4), pp. 363-375, 2001.

[12] X. Lin and P. Eades, “Towards area requirements for drawing hierarchically planar graphs”, Theoretical Computer Science, Volume 292, Issue 3, Pages 679-695, 31, 2003.

[13] J. Abello, “Hierarchical graph maps”, Computers & Graphics, Volume 28, Issue 3, Pages 345-359, June 2004.

[14] M. J¨unger and S. Leipert, "Level Planar Embedding", Journal of Graph Algorithms and Applications, vol. 6, no. 1, pp. 67-113, 2002.

[15] S. Liepert, “Level Planarity Testing an Embedding in Linear Time”, Ph.D. thesis, Unversität zu Köln ,1998.

[16] M. Jünger, S. Liepert and P. Mutzel, “Level Planarity Testing in Linear Time” , In S. Witesides, editor, Graph Drawing '98, volume 1547 of Lecture Notes in Computer Science, Springer Verlage, 224-237, 1998.

A Block Cipher Involving a Key Bunch Matrix and an Additional Key Matrix, Supplemented with Modular Arithmetic Addition and supported by Key-based Substitution

Dr. V.U.K.Sastry
Professor (CSE Dept), Dean (R&D)
SreeNidhi Institute of Science & Technology, SNIST
Hyderabad, India

K. Shirisha
Computer Science & Engineering
SreeNidhi Institute of Science & Technology, SNIST
Hyderabad, India

Abstract— **In this paper, we have devoted our attention to the development of a block cipher, which involves a key bunch matrix, an additional matrix, and a key matrix utilized in the development of a pair of functions called Permute() and Substitute(). These two functions are used for the creation of confusion and diffusion for each round of the iteration process of the encryption algorithm. The avalanche effect shows the strength of the cipher, and the cryptanalysis ensures that this cipher cannot be broken by any cryptanalytic attack generally available in the literature of cryptography.**

Keywords-key bunch matrix; additional key matrix; multiplicative inverse; encryption; decryption; permute; substitute.

I. Introduction

Security of information, which has to be maintained in a secret manner, is the primary concern of all the block ciphers. In a recent development, we have studied several block ciphers [1][2][3], "in press" [4], "unpublished" [5][6], "in press" [7], "unpublished" [8], wherein we have included a key bunch matrix and made use of the iteration process as a fundamental tool. In [7] and [8], we have introduced a key-based permutation and a key-based substitution for strengthen the cipher. Especially in [8], we have introduced an additional key matrix, supplemented with xor operation for adding some more strength to the cipher.

In the present paper, our objective is to modify the block cipher, presented in [7], by including and an additional key matrix supplemented with modular arithmetic addition. Here, our interest is to see how the permutation, the substitution and the additional key matrix would act in strengthening the cipher.

Now, let us mention the plan of the paper. We put forth the development of the cipher in section 2. Here, we portray the flowcharts and present the algorithms required in the development of this cipher. Then, we discuss the basic concepts of the key based permutation and substitution. We give an illustration of the cipher and discuss the avalanche effect, in section 3. We analyze the cryptanalysis, in section 4. Finally, we talk about the computations carried out in this analysis, and arrive at the conclusions, in section 5.

II. Developemnt Of The Cipher

Consider a plaintext matrix P, given by

$$P = [\, p_{ij} \,],\ i=1 \text{ to } n,\ j=1 \text{ to } n. \tag{2.1}$$

Let us take the key bunch matrix E in the form

$$E = [\, e_{ij} \,],\ i=1 \text{ to } n,\ j=1 \text{ to } n. \tag{2.2}$$

Here, we take all e_{ij} as odd numbers, which lie in the interval [1-255]. On using the concept of the multiplicative inverse, we get the decryption key bunch matrix D, in the form

$$D= [\, d_{ij} \,],\ i=1 \text{ to } n,\ j=1 \text{ to } n, \tag{2.3}$$

wherein d_{ij} and e_{ij} are related by the relation

$$(\, e_{ij} \times d_{ij} \,) \bmod 256 = 1, \tag{2.4}$$

for all i and j.

Here, it is to be noted that d_{ij} will be obtained as odd numbers and lie in the interval [1-255].

The additional key matrix F, can be taken in the form

$$F=[\, f_{ij} \,],\ i=1 \text{ to } n,\ j=1 \text{ to } n, \tag{2.5}$$

where f_{ij} are integers lying in [0-255].

The basic equations governing the encryption and the decryption, in this analysis, are given by

$$C = [\, c_{ij} \,] = (([\, e_{ij} \times p_{ij} \,] \bmod 256) + F) \bmod 256,$$
$$i=1 \text{ to } n,\ j = 1 \text{ to } n, \tag{2.6}$$

and

$$P = [\, p_{ij} \,] = [\, d_{ij} \times (C\text{-}F)_{ij} \,] \bmod 256,$$
$$i=1 \text{ to } n,\ j = 1 \text{ to } n, \tag{2.7}$$

where C is the ciphertext.

The flowcharts concerned to the procedure involved in this analysis are given in Figs. 1 and 2.

Here r denotes the number of rounds in the iteration process. The functions Permute() and Substitute() are used for achieving transformation of the plaintext, so that confusion and diffusion are created, in each round of the iteration process. The function Mult() is used to find the decryption key bunch matrix D from the given encryption key bunch matrix E. The functions IPermute() and ISubstitute() stand for the reverse process of the Permute() and Substitute(). The details of the permutation and substitution process are explained later.

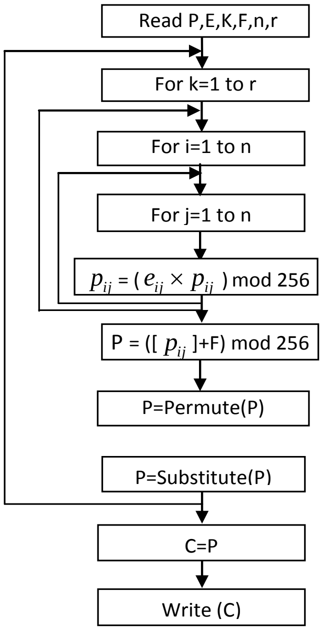


Figure 1 Flowchart for Encryption

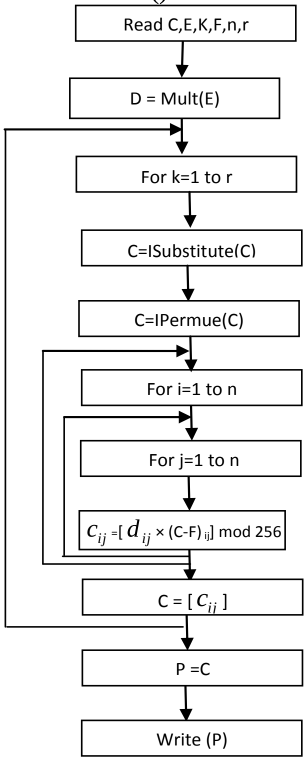


Figure 2 Flowchart for Decryption

The algorithms corresponding to the flowcharts are written as follows.

Algorithm for Encryption

```
1. Read P,E,K,F,n,r
2. For k = 1 to r do
   {
3.   For i=1 to n do
     {
4.     For j=1 to n do
       {
5.       $p_{ij}$ = ( $e_{ij}$ × $p_{ij}$ ) mod 256
       }
     }
6.   P=([ $p_{ij}$ ] + F) mod 256
7.   P=Permute(P)
8.   P=Substitute(P)
   }
8. C=P
9. Write(C)
```

Algorithm for Decryption

```
1. Read C,E,K,F,n,r
2. D=Mult(E)
3. For k = 1 to r do
   {
4.   C=ISubstitute(C)
5.   C=IPermute(C)
6.   For i =1 to n do
     {
7.     For j=1 to n do
       {
8.       $c_{ij}$ =[ $d_{ij}$ ×( $c_{ij}$ - $f_{ij}$ )] mod 256
       }
     }
9.   C=[ $c_{ij}$ ]
   }
10. P=C
11. Write (P)
```

In the development of the permutation and the substitution, we take a key matrix K in the form given below.

$$K = \begin{bmatrix} 156 & 14 & 33 & 96 \\ 253 & 107 & 110 & 127 \\ 164 & 10 & 5 & 123 \\ 174 & 202 & 150 & 94 \end{bmatrix} \qquad (2.8)$$

Figure 1. Flowchart for Encryption

The serial order, the elements in the key, the order of elements can be used and form a table of the form.

TABLE I. Relation Between Serial Numbers And Numbers In Ascending Order

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
156	14	33	96	253	107	110	127	164	10	5	123	174	202	150	94
12	3	4	6	16	7	8	10	13	2	1	9	14	15	11	5

In the process of permutation, we convert the decimal numbers in the plaintext matrix into binary bits and swap the rows firstly and the columns nextly, one after another, and achieve the final form of the permuted matrix by representing the binary bits in terms of decimal numbers. In the case of the substitution process, we consider the EBCDIC code matrix consisting of the decimal numbers 0 to 255, in 16 rows 16 columns, and interchange the rows firstly and the columns nextly, and then achieve the substitution matrix. For a detailed discussion of the functions Permute() and Substitute(), we refer to [7].

III. Illustration Of The Cipher And The Avalanche Effect

Consider the plaintext given below.

Dear Brother! I have got posting in army as a Captain a few days back. Both father and mother are advising me not to go there. They say that they have committed a sin in sending you as an Army Doctor. You know all the problems which you are facing in that environment in Indian Army. Tell me what shall I do? Would you suggest me to join in the same profession in which you are? All the retired Army employees who are residing in our area are telling “Serving Mother India is really great”. But most of their sons are working here only in our city. (3.1)

Let us focus our attention on the first 16 characters of the aforementioned plaintext. Thus we have

Dear Brother! I (3.2)

On using the EBCDIC code, the plaintext (3.2), can be written in the form P, given by

$$P = \begin{bmatrix} 196 & 133 & 129 & 153 \\ 64 & 194 & 153 & 150 \\ 163 & 136 & 133 & 153 \\ 79 & 64 & 201 & 64 \end{bmatrix}. \qquad (3.3)$$

Let us choose the encryption key bunch matrix E in the form

$$E = \begin{bmatrix} 9 & 81 & 201 & 137 \\ 235 & 93 & 15 & 107 \\ 33 & 79 & 191 & 255 \\ 57 & 197 & 179 & 3 \end{bmatrix}. \qquad (3.4)$$

We take the additional key matrix F in the form

$$F = \begin{bmatrix} 78 & 43 & 224 & 209 \\ 45 & 53 & 80 & 100 \\ 14 & 6 & 236 & 1 \\ 69 & 42 & 53 & 250 \end{bmatrix}. \qquad (3.5)$$

On using the concept of multiplicative inverse, mentioned in section 2, we get the decryption key bunch matrix D in the form

$$D = \begin{bmatrix} 57 & 177 & 121 & 185 \\ 195 & 245 & 239 & 67 \\ 225 & 175 & 63 & 255 \\ 9 & 13 & 123 & 171 \end{bmatrix}. \qquad (3.6)$$

On using the P, the E, and the F, given by (3.3) – (3.5), and applying the encryption algorithm, given in section 2, w get the ciphertext C in the form

$$C = \begin{bmatrix} 133 & 110 & 122 & 68 \\ 33 & 174 & 239 & 98 \\ 221 & 102 & 191 & 248 \\ 100 & 184 & 169 & 21 \end{bmatrix}. \qquad (3.7)$$

On using the C, the D, and the F, and applying the decryption algorithm, we get back the original plaintext P, given by (3.3).

Let us now examine the avalanche effect. On replacing the 2nd row 2nd column element 194 of the plaintext P, given by (3.3), by 195, we get the modified plaintext, wherein a change of one binary bit is there. On using this modified plaintext, the E and F, given by (3.4) and (3.5), and applying the encryption algorithm, we get the corresponding ciphertext.

$$C = \begin{bmatrix} 51 & 177 & 198 & 26 \\ 237 & 197 & 30 & 206 \\ 19 & 39 & 165 & 214 \\ 154 & 191 & 6 & 19 \end{bmatrix}. \qquad (3.8)$$

On comparing (3.7) and (3.8), after representing them in their binary form, we notice that these two ciphertexts differ by 72 bits out of 128 bits.

In a similar manner, let us offer one binary bit change in the encryption key bunch matrix E. This is achieved by replacing 3rd row 1st column element 33 of E by 32. Then on using this E, the original P, given by (3.3), the F, given by (3.5), and using the encryption algorithm, we obtain the corresponding ciphertext in the form

$$C = \begin{bmatrix} 155 & 158 & 195 & 250 \\ 156 & 158 & 6 & 221 \\ 151 & 186 & 1 & 19 \\ 127 & 39 & 20 & 221 \end{bmatrix}. \qquad (3.9)$$

On carrying out a comparative study of (3.7) and (3.9), after putting them in their binary form, we find that these two differ by 78 bits out of 128 bits. From the above discussion, we conclude that this cipher is exhibiting a strong avalanche effect, and the strength of the cipher is expected to be a remarkable one.

IV. Cryptanalysis

In the development of all the block ciphers, the importance of cryptanalysis is commendable. The different cryptanalytic attacks that are dealt with very often in the literature are

1. Ciphertext only attack (Brute force attack),
2. Known plaintext attack,
3. Chosen plaintext attack, and
4. Chosen ciphertext attack.

Generally, the first two attacks are examined in an analytical manner, while the latter two attacks are inspected with all care, in an intuitive manner. It is to be noted here that no cipher can be accepted, unless it withstands the first two attacks [9], and no cipher can be relied upon unless a clear cut decision is arrived in the case of the latter two attacks.

Let us now consider the brute force attack. In this analysis, we have 3 important entities namely, the key bunch matrix E, the additional key matrix F, and the special key K, used in the Permute() and Substitute() functions. On account of these three, the size of the key space can be written in the form

$$2^{7n^2} \times 2^{8n^2} \times 2^{128} = 2^{7n^2+8n^2+128} = 2^{15n^2+128}$$
$$= \left(2^{10}\right)^{\left(1.5n^2+12.8\right)} \approx \left(10^{3}\right)^{\left(1.5n^2+12.8\right)} = 10^{4.5n^2+38.4}$$

On assuming that, we require 10^{-7} seconds for computation with one set of keys in the key space, the time required for execution with all such possible sets in the key space is

$$\frac{10^{4.5n^2+38.4} \times 10^{-7}}{365 \times 24 \times 60 \times 60} = 3.12 \times 10^{4.5n^2+23.4}\ years.$$

In this analysis, as we have taken n=4, the time for computation with all possible sets of keys in the key space is

$$3.12 \times 10^{95.4} \, years.$$

As this is a very long span, this cipher cannot be broken by the brute force attack.

Now, let us examine the known plaintext attack. In the case of this attack, we know any number of plaintext and ciphertext pairs, which we require for our investigation. Focusing our attention on r=1, that is on the first round of the iteration process, in the encryption, we get the set of equations, given by

$$P=(([\, e_{ij} \times p_{ij} \,] \bmod 256)+F) \bmod 256, \ i=1 \text{ to } n, \ j=1 \text{ to } n, \quad (4.1)$$

$$P = \text{Permute}(P), \quad (4.2)$$

$$P = \text{Substitute}(P), \quad (4.3)$$

and

$$C = P \quad (4.4)$$

Here as C in (4.4) is known, we get P. However, as the substitution process and permutation process depend upon the key, one cannot have any idea regarding ISubstitute() and IPermute(). Thus it is simply impossible to determine P even at the next higher level that is in (4.3). In a spectacular manner, if one has a chance to know the key K (a rare situation), then one can determine P, occurring on the left hand side of (4.1), by using ISubstitute() and IPermute(). Then also, it is not at all possible to determine the e_{ij} (elements of the key bunch matrix), as this equation is totally involved on account of the presence of F and the mod operation. This shows that the cipher is strengthened by the presence of F.

From the above analysis, we conclude that this cipher cannot be broken by the known plaintext attack. As there are 16 rounds of iteration process, we can say very emphatically, that this cipher is unbreakable by the known plaintext attack.

On considering the set of equations in the encryption process, including mod, permute and substitute, we do not envisage any possible choice, either for the plaintext or for the ciphertext to make an attempt for breaking this cipher.

In the light of all these factors, we conclude that this cipher is a strong one and it can be applied for the secure transmission of any secret information.

V. Computations And Conclusions

In this paper, we have developed a block cipher which involves an encryption key bunch matrix, an additional matrix and a key matrix utilized for the development of a pair of functions called Permute() and Substitute(). In this analysis the additional matrix is supplemented with modular arithmetic addition. The cryptanalysis carried out in this investigation firmly indicates that this cipher cannot be broken by any cryptanalytic attack.

The programs required for encryption and decryption are written in Java.

The entire plain text given by (3.1) is divided into 3 blocks, wherein each block is written as a square matrix of size 16. As the last block is containing 37 characters, 219 zeroes are appended as additional characters so that it becomes a complete block.

To carry out the encryption of these plaintext blocks, here we take a key bunch matrix EK of size 16x16 and an additional matrix FK of the same size. They are taken in the form

$$EK = \begin{bmatrix} 19 & 173 & 1 & 247 & 187 & 205 & 221 & 157 & 129 & 15 & 249 & 125 & 69 & 127 & 193 & 245 \\ 149 & 35 & 205 & 117 & 177 & 15 & 161 & 173 & 51 & 185 & 203 & 61 & 79 & 93 & 239 & 33 \\ 211 & 213 & 207 & 29 & 91 & 237 & 159 & 9 & 49 & 29 & 69 & 35 & 113 & 49 & 179 & 119 \\ 161 & 147 & 77 & 53 & 67 & 169 & 203 & 189 & 159 & 113 & 185 & 181 & 59 & 19 & 117 & 43 \\ 65 & 221 & 195 & 171 & 145 & 253 & 65 & 115 & 229 & 173 & 147 & 63 & 181 & 147 & 11 & 109 \\ 179 & 119 & 53 & 45 & 11 & 205 & 97 & 145 & 223 & 135 & 239 & 21 & 155 & 83 & 133 & 183 \\ 7 & 45 & 71 & 177 & 57 & 203 & 145 & 189 & 221 & 191 & 197 & 109 & 227 & 131 & 1 & 75 \\ 153 & 103 & 119 & 209 & 43 & 189 & 149 & 67 & 243 & 155 & 95 & 39 & 117 & 67 & 251 & 135 \\ 181 & 157 & 185 & 11 & 153 & 127 & 55 & 241 & 73 & 205 & 255 & 227 & 229 & 149 & 9 & 21 \\ 187 & 203 & 159 & 107 & 91 & 197 & 229 & 37 & 177 & 23 & 205 & 153 & 177 & 93 & 253 & 241 \\ 239 & 115 & 233 & 187 & 227 & 71 & 85 & 249 & 175 & 77 & 29 & 245 & 69 & 179 & 189 & 249 \\ 17 & 197 & 27 & 45 & 141 & 117 & 161 & 91 & 191 & 145 & 45 & 229 & 49 & 145 & 191 & 77 \\ 107 & 105 & 245 & 75 & 99 & 185 & 97 & 211 & 151 & 239 & 229 & 105 & 233 & 155 & 179 & 213 \\ 247 & 221 & 111 & 231 & 135 & 209 & 181 & 251 & 85 & 37 & 119 & 91 & 93 & 93 & 15 & 221 \\ 157 & 89 & 199 & 121 & 193 & 23 & 47 & 115 & 159 & 127 & 203 & 167 & 3 & 239 & 249 & 47 \\ 141 & 191 & 103 & 107 & 221 & 251 & 79 & 147 & 249 & 41 & 91 & 225 & 177 & 85 & 5 & 155 \end{bmatrix}$$

and

$$FK = \begin{bmatrix} 58 & 125 & 140 & 75 & 9 & 209 & 148 & 230 & 62 & 52 & 94 & 184 & 76 & 195 & 213 & 28 \\ 190 & 223 & 33 & 102 & 237 & 11 & 93 & 234 & 147 & 163 & 125 & 171 & 56 & 7 & 47 & 123 \\ 141 & 52 & 198 & 148 & 83 & 159 & 15 & 128 & 0 & 169 & 193 & 116 & 114 & 232 & 167 & 32 \\ 26 & 0 & 245 & 81 & 199 & 230 & 79 & 190 & 222 & 197 & 202 & 169 & 8 & 10 & 241 & 47 \\ 189 & 148 & 30 & 85 & 174 & 52 & 195 & 76 & 33 & 100 & 35 & 141 & 109 & 73 & 205 & 244 \\ 110 & 197 & 159 & 67 & 112 & 191 & 126 & 234 & 66 & 138 & 239 & 108 & 98 & 148 & 188 & 40 \\ 1 & 146 & 84 & 215 & 77 & 151 & 44 & 141 & 238 & 148 & 120 & 182 & 208 & 20 & 182 & 5 \\ 100 & 50 & 54 & 3 & 76 & 29 & 103 & 143 & 241 & 174 & 1 & 75 & 240 & 32 & 70 & 187 \\ 92 & 10 & 136 & 150 & 207 & 134 & 188 & 135 & 231 & 109 & 108 & 134 & 103 & 115 & 153 & 188 \\ 70 & 15 & 26 & 201 & 69 & 242 & 229 & 42 & 43 & 19 & 55 & 129 & 178 & 47 & 255 & 96 \\ 85 & 8 & 25 & 80 & 129 & 120 & 182 & 205 & 135 & 249 & 68 & 12 & 131 & 41 & 98 & 95 \\ 212 & 70 & 239 & 99 & 44 & 204 & 49 & 3 & 38 & 173 & 243 & 228 & 111 & 252 & 32 & 174 \\ 233 & 62 & 187 & 61 & 221 & 230 & 87 & 203 & 71 & 39 & 16 & 160 & 139 & 105 & 232 & 41 \\ 88 & 135 & 212 & 153 & 82 & 54 & 35 & 220 & 49 & 185 & 13 & 214 & 97 & 120 & 251 & 155 \\ 197 & 205 & 217 & 159 & 69 & 217 & 54 & 143 & 232 & 27 & 19 & 252 & 202 & 238 & 96 & 166 \\ 253 & 35 & 224 & 212 & 105 & 100 & 184 & 216 & 31 & 40 & 93 & 125 & 38 & 127 & 145 & 244 \end{bmatrix}$$

On using each block of the plain text, the key bunch matrix EK and the additional matrix FK, in the places of E and F respectively, and applying the encryption algorithm, given in section 2, we carry out the encryption of each block separately, and obtain the cipher text as follows in (5.1).

Now, for the secure transmission of EK and FK, we encrypt these two by using E and F, and applying the encryption algorithm. Thus, we have the ciphertexts corresponding to EK and FK as given below, in (5.2) and (5.3), respectively.

From this analysis the sender transmits all the 3 blocks of the cipher text, corresponding to the entire plain text, and the cipher text of EK and FK, given in (5.1), (5.2) and (5.3), In addition to this information, he provides the key bunch matrix E, the additional matrix F and the key matrix K in a secured manner. He also supplies the number of characters with which the last block of the entire plain text is appended.

From the cryptanalysis carried out in this investigation we have found that this cipher is a strong one and cannot be broken by any cryptanalytic approach.

Here it may be noted that this cipher can be applied for the encryption of a plain text of any size, and for the encryption of a gray level or color image.

REFERENCES

[1] Dr. V.U.K. Sastry, K.Shirisha, "A Novel Block Cipher Involving a Key Bunch Matrix", in International Journal of Computer Applications (0975 – 8887) Volume 55– No.16, Oct 2012, Foundation of Computer Science, NewYork, pp. 1-6.

[2] Dr. V.U.K. Sastry, K.Shirisha, "A Block Cipher Involving a Key Bunch Matrix and Including Another Key Matrix Supplemented with Xor Operation ", in International Journal of Computer Applications (0975 – 8887) Volume 55– No.16, Oct 2012, Foundation of Computer Science, NewYork, pp.7-10.

[3] Dr. V.U.K. Sastry, K.Shirisha, "A Block Cipher Involving a Key Bunch Matrix and Including another Key Matrix Supported With Modular Arithmetic Addition", in International Journal of Computer Applications (0975 – 8887) Volume 55– No.16, Oct 2012, Foundation of Computer Science, NewYork, pp. 11-14.

[4] Dr. V.U.K. Sastry, K.Shirisha, "A novel block cipher involving a key bunch matrix and a permutation", International Journal of Computers and Electronics Research (IJCER), in press.

[5] Dr. V.U.K. Sastry, K.Shirisha, "A block cipher involving a key bunch matrix, and a key matrix supported with xor operation, and supplemented with permutation", unpublished.

[6] Dr. V.U.K. Sastry, K.Shirisha, "A block cipher involving a key bunch matrix, and a key matrix supported with modular arithmetic addition, and supplemented with permutation", unpublished.

[7] Dr. V.U.K. Sastry, K.Shirisha, "A novel block cipher involving a key bunch matrix and a key-based permutation and substitution", International Journal of Advanced Computer Science and Applications (IJACSA), in press.

[8] Dr. V.U.K. Sastry, K.Shirisha, "A block cipher involving a key bunch matrix and an additional key matrix, supplemented with xor operation and supported by key-based permutation and substitution", unpublished.

[9] William Stallings: Cryptography and Network Security: Principle and Practices", Third Edition 2003, Chapter 2, pp. 29.

AUTHORS PROFILE

Dr. V. U. K. Sastry is presently working as Professor in the Dept. of Computer Science and Engineering (CSE), Director (SCSI), Dean (R & D), SreeNidhi Institute of Science and Technology (SNIST), Hyderabad, India. He was Formerly Professor in IIT, Kharagpur, India and worked in IIT, Kharagpur during 1963 – 1998. He guided 14 PhDs, and published more than 86 research papers in various International Journals. He received the Best Engineering College Faculty Award in Computer Science and Engineering for the year 2008 from the Indian Society for Technical Education (AP Chapter), Best Teacher Award by Lions Clubs International, Hyderabad Elite, in 2012, and Cognizant-Sreenidhi Best faculty award for the year 2012. His research interests are Network Security & Cryptography, Image Processing, Data Mining and Genetic Algorithms.

K. Shirisha is currently working as Associate Professor in the Department of Computer Science and Engineering (CSE), SreeNidhi Institute of Science & Technology (SNIST), Hyderabad, India, since February 2007. She is pursuing her Ph.D. Her research interests are Information Security and Data Mining. She published three research papers in International Journals. She stood University topper in the M.Tech.(CSE).

55	98	195	171	226	83	253	114	163	77	26	121	39	199	193	190
164	159	13	81	117	43	52	60	154	205	38	146	142	105	68	54
207	35	52	107	192	208	193	24	11	134	252	36	22	193	196	242
137	191	244	80	8	206	7	54	132	31	41	140	41	117	208	75
203	252	146	129	11	160	217	143	120	11	71	59	233	193	72	157
207	205	10	87	46	13	213	20	189	137	189	135	141	161	228	145
128	199	218	65	87	12	184	133	242	130	101	119	97	88	92	183
193	97	33	94	174	219	138	41	37	96	60	23	76	21	185	251
229	141	212	251	102	227	180	135	49	137	134	100	181	60	106	198
66	82	216	84	228	85	204	106	178	97	12	240	173	186	55	241
123	221	164	106	78	109	157	41	7	23	32	69	251	59	236	231
33	203	137	21	213	28	83	187	20	74	53	108	190	234	125	5
24	10	66	74	123	3	105	179	41	164	100	79	23	21	31	128
251	239	115	49	124	75	19	28	41	72	105	187	36	70	205	92
49	29	162	253	39	251	109	65	118	18	254	252	159	94	120	123
195	238	106	186	180	251	183	37	245	173	112	16	5	231	2	236
187	166	0	84	24	113	0	176	211	250	131	95	63	67	84	164
134	204	147	101	67	157	191	24	236	80	159	245	130	60	185	171
228	33	237	209	121	30	14	243	202	80	147	109	247	83	39	170
118	64	144	24	233	138	109	121	90	68	110	4	242	220	207	239
216	45	26	38	1	226	4	25	174	6	239	164	185	103	71	121
47	207	34	153	29	125	155	186	228	219	192	226	45	120	154	50
98	128	235	220	8	40	163	154	164	67	103	115	129	148	90	85
67	181	251	59	120	53	97	7	37	210	192	15	33	252	84	152
109	128	185	230	65	141	198	227	119	64	247	106	151	163	5	8
150	166	129	130	17	54	1	38	180	69	36	15	102	78	106	134
14	200	51	243	192	162	200	43	64	52	90	16	1	70	193	34
126	78	156	252	57	84	199	200	29	104	46	101	151	0	96	111
225	152	219	108	60	187	22	161	75	205	76	206	117	216	3	199
57	200	162	99	52	22	205	88	75	61	141	183	72	235	174	7
172	232	228	31	240	105	180	85	207	189	252	134	77	144	148	141
248	27	132	35	154	195	161	209	176	169	136	78	229	160	180	79
244	161	218	39	227	184	49	171	105	36	203	137	166	210	242	135
135	58	61	235	246	199	126	224	136	164	228	42	229	34	204	252
161	231	179	113	141	146	197	197	243	230	188	69	60	148	23	42
14	109	166	239	54	23	117	182	67	7	52	83	113	219	42	163
137	74	198	183	247	73	133	93	205	23	19	61	1	63	61	155
59	66	89	105	102	217	107	74	169	72	72	98	140	196	253	2
34	178	246	157	240	116	218	205	49	207	44	185	190	252	50	180
29	34	126	43	89	96	100	149	233	132	102	192	48	51	25	154
190	34	18	109	217	108	90	205	64	145	113	70	54	138	191	29
160	157	192	74	218	189	99	89	68	125	239	199	24	216	22	21
255	198	147	22	53	89	164	99	93	146	233	217	219	121	212	61
231	38	174	103	125	63	175	178	147	30	9	175	197	167	200	177
197	85	90	248	190	225	96	74	45	19	35	194	157	158	198	31
233	108	66	0	56	114	65	50	87	15	205	89	91	80	241	146
85	132	187	63	151	245	175	211	114	121	31	155	199	186	229	116
183	64	216	127	196	21	229	173	252	71	135	143	85	245	162	78

(5.1)

116	112	40	123	211	102	93	179	40	154	235	69	34	147	243	36
146	180	23	213	21	186	167	12	57	85	65	84	121	78	180	31
224	176	75	84	49	185	144	147	170	205	61	200	217	72	100	207
105	110	246	250	158	251	111	164	49	10	62	52	231	245	237	106
90	72	239	74	160	4	183	54	28	243	51	135	161	194	153	80
251	35	250	13	222	66	16	246	78	20	98	115	121	242	111	239
13	94	140	164	189	182	31	5	42	244	230	117	228	231	67	239
101	190	72	68	226	46	188	215	238	127	152	114	121	99	19	10
155	224	45	11	206	8	98	81	126	233	95	3	166	44	133	97
161	116	250	217	241	169	79	197	219	216	182	98	160	100	24	127
131	51	198	162	250	246	201	116	118	76	160	124	72	132	38	50
144	170	99	186	250	165	87	62	147	19	114	104	131	14	204	188
191	160	18	37	247	233	129	220	199	40	71	96	171	108	253	92
129	101	41	89	89	4	247	147	144	12	4	122	210	78	249	103
42	10	255	126	157	148	99	255	173	214	52	200	113	215	190	231
181	131	98	6	241	203	213	96	64	95	99	135	253	228	136	213

(5.2)

and

58	125	140	75	9	209	148	230	62	52	94	184	76	195	213	28
190	223	33	102	237	11	93	234	147	163	125	171	56	7	47	123
141	52	198	148	83	159	15	128	0	169	193	116	114	232	167	32
26	0	245	81	199	230	79	190	222	197	202	169	8	10	241	47
189	148	30	85	174	52	195	76	33	100	35	141	109	73	205	244
110	197	159	67	112	191	126	234	66	138	239	108	98	148	188	40
1	146	84	215	77	151	44	141	238	148	120	182	208	20	182	5
100	50	54	3	76	29	103	143	241	174	1	75	240	32	70	187
92	10	136	150	207	134	188	135	231	109	108	134	103	115	153	188
70	15	26	201	69	242	229	42	43	19	55	129	178	47	255	96
85	8	25	80	129	120	182	205	135	249	68	12	131	41	98	95
212	70	239	99	44	204	49	3	38	173	243	228	111	252	32	174
233	62	187	61	221	230	87	203	71	39	16	160	139	105	232	41
88	135	212	153	82	54	35	220	49	185	13	214	97	120	251	155
197	205	217	159	69	217	54	143	232	27	19	252	202	238	96	166
253	35	224	212	105	100	184	216	31	40	93	125	38	127	145	244

(5.3)

A Novel Block Cipher Involving a Key bunch Matrix and a Key-based Permutation and Substitution

Dr. V.U.K.Sastry
Professor (CSE Dept), Dean (R&D)
SreeNidhi Institute of Science & Technology, SNIST
Hyderabad, India

K. Shirisha
Computer Science & Engineering
SreeNidhi Institute of Science & Technology, SNIST
Hyderabad, India

Abstract— **In this paper, we have developed a novel block cipher involving a key bunch matrix supported by a key-based permutation and a key-based substitution. In this analysis, the decryption key bunch matrix is obtained by using the given encryption key bunch matrix and the concept of multiplicative inverse. From the cryptanalysis carried out in this investigation, we have seen that the strength of the cipher is remarkably good and it cannot be broken by any conventional attack.**

Keywords- Key bunch matrix; encryption; decryption; permutation; substitution; avalanche effect; cryptanalysis.

I. INTRODUCTION

The development of block ciphers, basing upon a secret key, is a fascinating area of research in cryptography. Though there are several block ciphers, such as Hill Cipher [1], Fiestal Cipher [2], DES [3], together with its variants [4][5], and AES [6]. In all these ciphers, the processes, namely, iteration, permutation and substitution play a vital role in strengthening the cipher. More often, in all these ciphers, the block length and the key length are maintained as 64, 128, 192, or 256 binary bits.

In a recent investigation, we have developed a set of block ciphers [7], [8], [9], "in press" [10], "unpublished" [11], [12], wherein, a secret key bunch matrix plays a prominent role. In all these ciphers, the encryption key bunch matrix contains a set of keys, in which each key is an odd number lying in [1-255]. In all these analyses, the corresponding decryption key bunch matrix, which is also containing odd numbers lying in [1-255], is obtained by using the concept of the multiplicative inverse [4]. In the development of all these block ciphers, the length of the plaintext can be taken as large as possible, at our will, as the size of the key bunch matrix can be chosen as big as possible, in an effective manner. This feature ensures the strength of the cipher in a remarkable way.

In the present investigation, our objective is to develop a novel block cipher, by using the encryption key bunch matrix, and applying a key-based permutation and substitution which strengthen the cipher in a significant manner. The details of the permutation and the substitution processes are presented later.

In what follows, we mention the plan of the paper. In section 2, we discuss the development of the cipher. Further, we present flowcharts and algorithms required in this investigation. Here we deal with the key based permutation and substitution involved in this analysis. In section 3, we offer an illustration of the cipher. In this, we examine the avalanche effect, which acts as a benchmark in respect of the strength of the cipher. In section 4, we make a study of the cryptanalysis. Finally in section 5, we present the computations carried out in this analysis, and arrive at conclusions.

II. DEVELOPMENT OF THE CIPHER

Consider a plaintext P which can be represented in the form of a matrix given by

$$P = [\ p_{ij}\ ],\ i=1 \text{ to } n,\ j=1 \text{ to } n, \qquad (2.1)$$

wherein each p_{ij} is a decimal number lying in [0-255].

Let

$$E = [\ e_{ij}\ ],\ i=1 \text{ to } n,\ j=1 \text{ to } n, \qquad (2.2)$$

be the encryption key bunch matrix, in which each e_{ij} is an odd number lying in [1-255], and

$$D= [\ d_{ij}\ ],\ i=1 \text{ to } n,\ j=1 \text{ to } n, \qquad (2.3)$$

be the decryption key bunch matrix, wherein each d_{ij} is an odd number lying in [1-255]. e_{ij} and d_{ij} are connected by the relation

$$(\ e_{ij} \times d_{ij}\ ) \bmod 256 = 1, \qquad (2.4)$$

Here it may be noted that the d_{ij} is obtained corresponding to every given e_{ij} in an appropriate manner.

The basic equations governing the encryption and the decryption processes of the cipher can be written in the form

$$C = [\ c_{ij}\ ]=[\ e_{ij} \times p_{ij}\ ] \bmod 256,\ i=1 \text{ to } n,\ j = 1 \text{ to } n \qquad (2.5)$$

and

$$P = [\ p_{ij}\ ]=[\ d_{ij} \times c_{ij}\ ] \bmod 256,\ i=1 \text{ to } n,\ j = 1 \text{ to } n. \qquad (2.6)$$

On assuming that the cipher involoves an iteration process, the flowcharts governing the encryption and the decryption can be drawn as shown in Figs. 1 and 2.

In this analysis, r denotes the number of rounds in the iteration process, and is taken as 16.

The function Substitute(), occurring in the flowchart of the encryption, denotes the key-dependant substitution process, that we are going to describe a little later. The function ISubstitute(), occurring in the decryption process, denotes the reverse process of the Substitute(). The function Mult(), which is in the decryption process, is used to find the decryption key bunch matrix D from the given encryption key bunch matrix E.

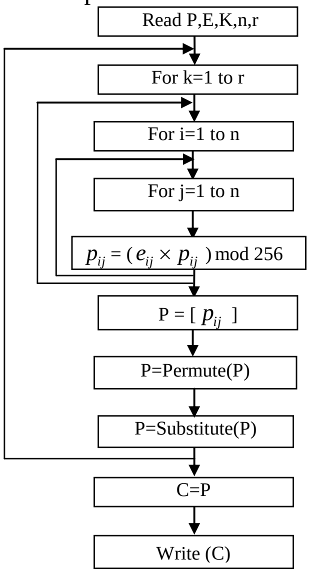


Figure 1. Flowchart for Encryption

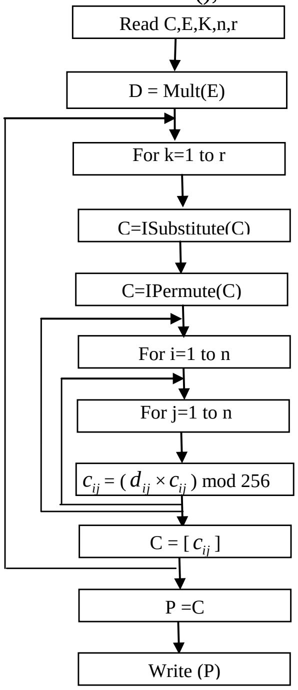


Figure 2. Flowchart for Decryption

The corresponding algorithms for the encryption and the decryption are written as follows.

Algorithm for Encryption

```
1. Read P,E,K,n,r
2. For k = 1 to r do
   {
3.    For i=1 to n do
      {
4.       For j=1 to n do
         {
5.          $p_{ij}$ = ( $e_{ij}$ × $p_{ij}$ ) mod 256
         }
      }
6.    P=[ $p_{ij}$ ]
7.    P=Permute(P)
8.    P=Substitute(P)
   }
8. C=P
9. Write(C)
```

Algorithm for Decryption

```
1. Read C,E,K,n,r
2. D=Mult(E)
3. For k = 1 to r do
   {
4.    C=ISubstitute(C)
5.    C=IPermute(C)
6.    For i =1 to n do
      {
7.       For j=1 to n do
         {
8.          $c_{ij}$ = ( $d_{ij}$ × $c_{ij}$ ) mod 256
         }
      }
9.    C=[ $c_{ij}$ ]
   }
10. P=C
11. Write (P)
```

To have a clear insight into the key dependent permutation process and key dependent substitution process, which we are adopting in this analysis, let us consider a typical example. Let us take a key K in the form

$$K = \begin{bmatrix} 156 & 14 & 33 & 96 \\ 253 & 107 & 110 & 127 \\ 164 & 10 & 5 & 123 \\ 174 & 202 & 150 & 94 \end{bmatrix} \qquad (2.7)$$

We write the elements of this key in a tabular form as shown below.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
156	14	33	96	253	107	110	127	164	10	5	123	174	202	150	94

Here the first row shows the serial number and the second row is concerned to the elements in the key K.

On considering the order of magnitude of the elements in the key, we can write the above table, by including one more row, in the following form

TABLE I. RELATION BETWEEN SERIAL NUMBERS AND NUMBERS IN ASCENDING ORDER

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
156	14	33	96	253	107	110	127	164	10	5	123	174	202	150	94
12	3	4	6	16	7	8	10	13	2	1	9	14	15	11	5

Here the 3rd row denotes the order of magnitude of the elements in the key.

The process of permutation, basing upon the key used in this analysis, can be explained as follows. Let

$$x_1, x_2, x_3, \ldots, x_{14}, x_{15}, x_{16}$$

be a set of numbers. On using the numbers, occurring in the first and third rows of the Table-1, we swap the pairs (x_1, x_{12}), (x_2, x_3), (x_4, x_6), (x_5, x_{16}), (x_7, x_8), (x_9, x_{13}) and (x_{14}, x_{15}). Here it is to be noted that, (x3, x4) are not swapped, as x3 is already swapped with x2. Similarly, we do not do any swapping in the case of the numbers (x_3, x_4), (x_6, x_7), (x_8, x_{10}), (x_{10}, x_2), (x_{11}, x_1), (x_{12}, x_9), (x_{13}, x_{14}), (x_{15}, x_{11}) and (x_{16}, x_5). This is the basic idea of the permutation process, which we employ in the case of columns of numbers as well as rows of numbers occurring in a matrix. For clarity of this process, we refer to the illustration that we are going to do in section 3, a little later.

Let us firstly discuss the process of the key based permutation applied on a plaintext obtained in any round of the iteration process of the encryption. Consider the plaintext P= [p_{ij}], i=1 to n, j=1 to n. Let us consider the first two rows of this matrix. On representing each decimal number p_{ij} in its binary form, and writing the binary bits in a vertical manner, we get a matrix of size 16xn, for these two rows. On assuming that n is divisible by 16 (for convenience), we can represent these two rows in the form of n/16 sub-matrices, wherein each one is a square matrix of size 16. Then on swapping the rows (as pointed out in the case of the numbers x1 to x16) and the columns (subsequently one after another), we get the corresponding permuted matrices. After that, by taking the binary bits in a row-wise manner, we convert them into decimal numbers, and write them in a row-wise manner. Thus we get back a matrix of size 2×n.We carry out this process in a similar manner for every pair of rows and having n columns. Thus we complete the permutation of the entire matrix and get a permuted matrix of size nxn. However if n<16, the process of swapping is restricted according to the value of n. For example, let us suppose that n=4. And P is of the form given by

$$P = \begin{bmatrix} 198 & 34 & 45 & 12 \\ 56 & 92 & 101 & 223 \\ 175 & 49 & 245 & 0 \\ 211 & 65 & 8 & 100 \end{bmatrix} \tag{2.8}$$

On writing the 16 decimal numbers in terms of binary bits in a column-wise manner, the matrix (2.8) can be represented in the form of a matrix of size 8x16. This is given by

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \end{bmatrix} \tag{2.9}$$

Firstly, as suggested by Table-1, we interchange the row pairs (2,3), (4,6), and (7,8). Thus we get

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{2.10}$$

We need not interchange rows any more as we have only 8 rows in this matrix. Now, we interchange the columns following the information in Table-1. This will lead to a matrix of size 8x16, which is given by

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{2.11}$$

This completes the process of the permutation, denoted by the function Permute().

Let us now describe the process of the key-based substitution. We now consider the numbers [0-255] that are occurring in EBCDIC table. These numbers can be represented in the form of a square matrix of size 16 by writing the table in the form

$$EB(i,j) = [16(i-1) + j - 1],\ i = 1\ to\ 16,\ j = 1\ to\ 16 \tag{2.12}$$

On using the basic idea of the key-based permutation process, we permute the rows (firstly) and the columns (subsequently), and obtain the substitution matrix, called SB, given by

$$SB = \begin{bmatrix} 187 & 178 & 177 & 181 & 191 & 179 & 183 & 182 & 188 & 185 & 186 & 176 & 184 & 190 & 189 & 180 \\ 43 & 34 & 33 & 37 & 47 & 35 & 39 & 38 & 44 & 41 & 42 & 32 & 40 & 46 & 45 & 36 \\ 27 & 18 & 17 & 21 & 31 & 19 & 23 & 22 & 28 & 25 & 26 & 16 & 24 & 30 & 29 & 20 \\ 91 & 82 & 81 & 85 & 95 & 83 & 87 & 86 & 92 & 89 & 90 & 80 & 88 & 94 & 93 & 84 \\ 251 & 242 & 241 & 245 & 255 & 243 & 247 & 246 & 252 & 249 & 250 & 240 & 248 & 254 & 253 & 244 \\ 59 & 50 & 49 & 53 & 63 & 51 & 55 & 54 & 60 & 57 & 58 & 48 & 56 & 62 & 61 & 52 \\ 123 & 114 & 113 & 117 & 127 & 115 & 119 & 118 & 124 & 121 & 122 & 112 & 120 & 126 & 125 & 116 \\ 107 & 98 & 97 & 101 & 111 & 99 & 103 & 102 & 108 & 105 & 106 & 96 & 104 & 110 & 109 & 100 \\ 203 & 194 & 193 & 197 & 207 & 195 & 199 & 198 & 204 & 201 & 202 & 192 & 200 & 206 & 205 & 196 \\ 155 & 146 & 145 & 149 & 159 & 147 & 151 & 150 & 156 & 153 & 154 & 144 & 152 & 158 & 157 & 148 \\ 171 & 162 & 161 & 165 & 175 & 163 & 167 & 166 & 172 & 169 & 170 & 160 & 168 & 174 & 173 & 164 \\ 11 & 2 & 1 & 5 & 15 & 3 & 7 & 6 & 12 & 9 & 10 & 0 & 8 & 14 & 13 & 4 \\ 139 & 130 & 129 & 133 & 143 & 131 & 135 & 134 & 140 & 137 & 138 & 128 & 136 & 142 & 141 & 132 \\ 235 & 226 & 225 & 229 & 239 & 227 & 231 & 230 & 236 & 233 & 234 & 224 & 232 & 238 & 237 & 228 \\ 219 & 210 & 209 & 213 & 223 & 211 & 215 & 214 & 220 & 217 & 218 & 208 & 216 & 222 & 221 & 212 \\ 75 & 66 & 65 & 69 & 79 & 67 & 71 & 70 & 76 & 73 & 74 & 64 & 72 & 78 & 77 & 68 \end{bmatrix} \tag{2.13}$$

The function Substitute() works as follows: On noticing the position of a decimal number (corresponding to a character in the plaintext, at any stage of the iteration process) in the EBCDIC table, we substitute that number in the plaintext by the decimal number occurring in the same position of the substitution matrix.

The functions IPermute() and ISubstitute() denote the reverse processes of the Permute() and the Substitute(), respectively. The function Mult() is used to find the decryption key bunch matrix D for the given encryption key bunch matrix E.

III. Illustration Of The Cipher And The Avalanche Effect

Consider the plaintext given below.

Dear Brother-in-law! Up to the time that you went abroad, that is a month back, my mother and father promised to give me to you in marriage. They do not want their daughter to go away to this country. They say that they cannot live without my presence along with this in this country. Now they are searching for an Indian match. You are highly qualified. You did your M.Tech. Now you are doing your Doctorate. How can I forget you? I all the while remember your charming personality and your pleasant talk. It is simply impossible for me to forget you and marry someone else. Whatever my father and mother say to me I want to escape from their clutches and reach you as early as possible. I am finishing my final year exams. I have already passed GRE and TOEFL. I would apply for bank loan with the cooperation of your father and get away from this country very soon and join you without any second thought. (3.1)

Let us focus our attention on the first 16 characters of the plaintext. This is given by

Dear Brother-in- (3.2)

On using the EBCDIC code, the plaintext (3.2) can be written in the form of a matrix P given by

$$P = \begin{bmatrix} 196 & 133 & 129 & 153 \\ 64 & 194 & 153 & 150 \\ 163 & 136 & 133 & 153 \\ 96 & 137 & 149 & 96 \end{bmatrix}. \qquad (3.3)$$

Let us take the encryption key bunch matrix E in the form

$$E = \begin{bmatrix} 21 & 57 & 171 & 39 \\ 101 & 67 & 89 & 223 \\ 67 & 157 & 171 & 1 \\ 37 & 203 & 233 & 17 \end{bmatrix}. \qquad (3.4)$$

On applying the concept of the multiplicative inverse, we get

$$D = \begin{bmatrix} 61 & 9 & 3 & 151 \\ 109 & 107 & 233 & 31 \\ 107 & 181 & 3 & 1 \\ 173 & 227 & 89 & 241 \end{bmatrix}. \qquad (3.5)$$

On using the plaintext P, the encryption key bunch matrix E and the encryption algorithm, given in section 2, we get the ciphertext C in the form

$$C = \begin{bmatrix} 20 & 197 & 152 & 47 \\ 247 & 232 & 171 & 142 \\ 91 & 154 & 73 & 113 \\ 168 & 34 & 170 & 80 \end{bmatrix}. \qquad (3.6)$$

Now, on using the decryption key bunch matrix D, given by (3.5), the ciphertext C, given by (3.6), and applying the decryption algorithm, we get back the plaintext P, given by (3.3).

Let us now examine the avalanche effect. On replacing the 4th row 2nd column element, 137 by 169, we get a change of one binary bit in the plaintext. On using this modified plaintext, the encryption key bunch matrix E and applying the encryption algorithm, we get a new ciphertext C in the form

$$C = \begin{bmatrix} 176 & 187 & 193 & 16 \\ 120 & 5 & 219 & 17 \\ 75 & 35 & 72 & 174 \\ 252 & 3 & 116 & 221 \end{bmatrix}. \qquad (3.7)$$

On comparing (3.6) and (3.7), after converting them binary form, we notice that these two ciphertexts differ by 68 bits out of 128 bits. Let us now consider the case of a one bit change in the key bunch matrix E. This can be achieved by replacing 101 (the 2nd row 1st column element of E) by 116. Now, on using the modified E, the plaintext P, given by (3.3), and applying the encryption algorithm, we get the corresponding ciphertext C in the form

$$C = \begin{bmatrix} 204 & 86 & 71 & 1 \\ 77 & 69 & 102 & 100 \\ 235 & 116 & 221 & 186 \\ 45 & 76 & 235 & 186 \end{bmatrix}. \qquad (3.8)$$

On converting the ciphertexts (3.6) and (3.8) into their binary form, and comparing them, we find that these two ciphertexts differ by 71 bits out of 128 bits.

From the above analysis, we conclude that the cipher is expected to be a strong one.

IV. Cryptanalysis

In the literature of the cryptography, the strength of a cipher can be decided by carrying out cryptanalysis. The different attacks that are available for breaking a cipher are

1. Ciphertext only attack (Brute force attack),
2. Known plaintext attack,
3. Chosen plaintext attack, and
4. Chosen ciphertext attack.

Generally every cipher is designed, so that it withstands the first two attacks [4]. However the latter two attacks are examined intuitively and checked up whether the cipher can be broken by those attacks.

Let us now consider the ciphertext only attack. In this cipher, the encryption key bunch matrix is of size n×n. The key matrix used in the development of the permutation and the substitution is a square matrix of size 4. Hence the size of the key space is

$$2^{7n^2+128} = (2^{10})^{0.7n^2+12.8} \approx 10^{2.1n^2+38.4}$$

If we assume that the time required for the computation of the cipher with one value of the key in the key space is 10^{-7} seconds, then the time required for the execution of the cipher with all possible values of the key in the key space is

$$\frac{10^{2.1n^2+38.4}\times 10^{-7}}{365\times 24\times 60\times 60}=\frac{10^{2.1n^2+31.4}}{365\times 24\times 60\times 60}$$

$$=3.12\times 10^{2.1n^2+23.4}\text{ years}$$

In this analysis, as we have taken n=4, the time required for the execution assumes the form $3.12\times 10^{33.6}$ years. As this is a very large number, it is simply impossible to break this cipher by the brute force attack.

Let us now consider the known plaintext attack. In order to carry out this one, we know as many pairs of plaintexts and ciphertexts as we require. If we confine our attention to r=1, that is to the first round of the iteration process, then the basic equations governing the cipher are given by

P = [$e_{ij}\times p_{ij}$] mod 256, i = 1 to n, j=1 to n, (4.1)

P = Permute(P), (4.2)

P = Substitute(P), (4.3)

and

C = P (4.4)

As C is known to us, the P on the right side of (4.4) is known. Thus, though P on the left side of (4.3) is known to us, the P on the right side of (4.3) cannot be determined as the Substitute() and the ISubstitute(), which depend upon the key K, are unknown to us. Hence this cipher cannot be broken by the known plaintext attack, even when r=1, as K is not known. However, if an attempt is made to tackle this problem by the brute force attack, that is choosing K in all possible ways, covering the entire key space of the key K, then the time required for developing the functions Permute() and Substitute() can be shown to be

$$\frac{2^{128}\times 10^{-7}}{365\times 24\times 60\times 60}=3.12\times 10^{23.4}\,years.$$

as the length of the key K is 128 binary bits. Here, it is assumed that the time required for the computation of Permute() and Substitute() (together with IPermute() and ISubstitute()) takes 10^{-7} seconds. As this time is very large, we firmly conclude that this cipher cannot be broken by the known plaintext attack, even when we supplement it with the brute force attack.

As the equations governing the cipher, are non-linear and highly involved, due to permutation, substitution and modular arithmetic operations, we envisage that it is not possible to choose either a plaintext or a ciphertext for breaking the cipher by the third or the fourth attack.

In the light of the above facts, we conclude that, this cipher is a strong one and it cannot be broken by any conventional attack.

V. COMPUTATIONS AND CONCLUSIONS

In this investigation, we have developed a novel block cipher by using a key bunch matrix. In this, we have made use of a permutation process and a substitution process basing upon a key matrix of size 4x4. The strength of a cipher has increased enormously as we have introduced iteration process and the functions Permute() and Substitute().

The programs required for encryption and decryption are written in Java.

When the size of the plaintext is very large, it is rather tedious to carry out the encryption process by using a key bunch matrix E of size 4x4. Thus, in order to carry out the encryption of the entire plaintext, given in (3.1), we take a key bunch matrix EK of size 16x16. This is taken in the form, given by (5.1).

$$EK=\begin{bmatrix}
49 & 163 & 109 & 217 & 133 & 161 & 225 & 89 & 163 & 209 & 225 & 255 & 39 & 31 & 235 & 169\\
13 & 227 & 207 & 107 & 207 & 67 & 191 & 161 & 143 & 215 & 29 & 179 & 133 & 45 & 57 & 5\\
253 & 211 & 79 & 121 & 91 & 95 & 167 & 89 & 157 & 159 & 111 & 175 & 249 & 71 & 213 & 139\\
233 & 195 & 247 & 7 & 231 & 185 & 41 & 243 & 223 & 81 & 83 & 113 & 149 & 27 & 1 & 213\\
91 & 129 & 73 & 47 & 187 & 245 & 115 & 143 & 153 & 209 & 31 & 27 & 243 & 39 & 159 & 11\\
131 & 185 & 23 & 17 & 187 & 255 & 169 & 97 & 55 & 157 & 149 & 199 & 247 & 85 & 61 & 27\\
255 & 209 & 29 & 95 & 77 & 183 & 117 & 145 & 107 & 139 & 91 & 1 & 227 & 87 & 243 & 9\\
133 & 93 & 49 & 111 & 115 & 131 & 239 & 63 & 141 & 137 & 193 & 23 & 45 & 193 & 179 & 217\\
217 & 97 & 19 & 245 & 113 & 83 & 103 & 159 & 147 & 49 & 225 & 41 & 247 & 193 & 99 & 139\\
151 & 143 & 191 & 205 & 91 & 151 & 197 & 137 & 23 & 151 & 103 & 91 & 109 & 91 & 11 & 65\\
249 & 39 & 33 & 143 & 69 & 247 & 243 & 53 & 11 & 211 & 99 & 119 & 13 & 19 & 207 & 221\\
223 & 101 & 225 & 233 & 61 & 111 & 201 & 149 & 3 & 1 & 55 & 121 & 3 & 175 & 101 & 91\\
85 & 61 & 95 & 195 & 33 & 41 & 33 & 71 & 151 & 43 & 93 & 233 & 193 & 159 & 13 & 97\\
175 & 93 & 9 & 99 & 59 & 73 & 167 & 127 & 247 & 95 & 135 & 203 & 29 & 55 & 25 & 163\\
231 & 215 & 131 & 237 & 131 & 93 & 255 & 181 & 211 & 107 & 77 & 47 & 91 & 249 & 39 & 105\\
75 & 225 & 189 & 41 & 75 & 251 & 193 & 79 & 199 & 101 & 95 & 179 & 63 & 189 & 67 & 19
\end{bmatrix}$$

(5.1)

The plaintext given in (3.1) is containing 907 characters. This can be divided into 4 blocks, wherein each block is containing 256 characters. However, we have appended 117 zeroes characters so that we make the last block a complete block. Now, on using K and EK, given in (2.7) and (5.1), and the encryption process, given in section 2, four times, we get the cipher text in the form, given in (5.2).

In order to send the size key bunch matrix EK, in a secret manner, let us encrypt this one by using E as the key bunch matrix. Thus we arrive at the ciphertext corresponding to EK as shown in (5.3).

It is to be noted here, that the sender has to send the ciphertext corresponding to entire plaintext, the number of characters added in the last block, and the ciphertext corresponding to EK to the receiver. Further the sender has to provide E and K in a secret manner.

From the above analysis, we notice that this cipher is a strong one and it can be applied for the transmission of a plaintext of any length in a secured manner. It may also be noted here that this cipher is very much useful in encrypting black and white images and color images.

REFERENCES

[1] Lester Hill, (1929), “Cryptography in an algebraic alphabet”, (V.36 (6), pp. 306-312.), American Mathematical Monthly.

[2] Fiestal H., Cryptography and Computer Privacy, Scientific American, May 1973.

[3] National Bureau of Standards NBS FIPS PUB 46 “Data Encryption Standard (DES)”, US Department of Commerce, January 1977.

[4] William Stallings: Cryptography and Network Security: Principle and Practices”, Third Edition 2003, Chapter 2, pp. 29.

[5] Tuchman, W., “ Hellman presents no Shortcut Solutions to DES”, IEEE Spectrum, July, 1979.

[6] Daemen J., Rijman V., “Rijndael, The Advanced Encryption Standard (AES)”, Dr. Dobb’s Journal, vol. 26, No. 3, March 2001, pp. 137-139.

[7] Dr. V.U.K. Sastry, K.Shirisha, “A Novel Block Cipher Involving a Key Bunch Matrix”, in International Journal of Computer Applications (0975 – 8887) Volume 55– No.16, Oct 2012, Foundation of Computer Science, NewYork, pp. 1-6.

[8] Dr. V.U.K. Sastry, K.Shirisha, “A Block Cipher Involving a Key Bunch Matrix and Including Another Key Matrix Supplemented with Xor Operation ”, in International Journal of Computer Applications (0975 – 8887) Volume 55– No.16, Oct 2012, Foundation of Computer Science, NewYork, pp.7-10.

Dr. V.U.K. Sastry, K.Shirisha, “A Block Cipher Involving a Key Bunch Matrix and Including another Key Matrix Supported With Modular Arithmetic Addition”, in International Journal of Computer Applications (0975 – 8887) Volume 55– No.16, Oct 2012, Foundation of Computer Science, NewYork, pp. 11-14.

[9] Dr. V.U.K. Sastry, K.Shirisha, “A novel block cipher involving a key bunch matrix and a permutation”, International Journal of Computers and Electronics Research (IJCER), in press.

[10] Dr. V.U.K. Sastry, K.Shirisha, “A block cipher involving a key bunch matrix, and a key matrix supported with xor operation, and supplemented with permutation”, unpublished.

[11] Dr. V.U.K. Sastry, K.Shirisha, “A block cipher involving a key bunch matrix, and a key matrix supported with modular arithmetic addition, and supplemented with permutation”, unpublished.

AUTHORS PROFILE

Dr. V. U. K. Sastry is presently working as Professor in the Dept. of Computer Science and Engineering (CSE), Director (SCSI), Dean (R & D), SreeNidhi Institute of Science and Technology (SNIST), Hyderabad, India. He was Formerly Professor in IIT, Kharagpur, India and worked in IIT, Kharagpur during 1963 – 1998. He guided 14 PhDs, and published more than 86 research papers in various International Journals. He received the Best Engineering College Faculty Award in Computer Science and Engineering for the year 2008 from the Indian Society for Technical Education (AP Chapter), Best Teacher Award by Lions Clubs International, Hyderabad Elite, in 2012, and Cognizant- Sreenidhi Best faculty award for the year 2012. His research interests are Network Security & Cryptography, Image Processing, Data Mining and Genetic Algorithms.

K. Shirisha is currently working as Associate Professor in the Department of Computer Science and Engineering (CSE), SreeNidhi Institute of Science & Technology (SNIST), Hyderabad, India, since February 2007. She is pursuing her Ph.D. Her research interests are Information Security and Data Mining. She published three research papers in International Journals. She stood University topper in the M.Tech.(CSE).

223	241	161	13	58	52	154	202	32	81	6	150	237	156	161	183
121	39	196	90	88	91	197	252	96	78	118	17	201	95	137	127
189	132	82	3	45	208	66	85	62	158	217	227	42	11	113	104
129	160	72	21	246	93	91	29	75	113	73	79	246	108	54	97
88	219	168	114	10	133	194	178	249	91	152	182	241	251	74	148
233	148	80	51	235	204	235	115	239	223	38	40	24	64	34	65
105	227	176	240	113	3	12	74	151	190	81	165	7	112	111	241
130	153	4	158	188	202	15	197	52	225	121	52	84	3	214	24
198	36	184	60	138	1	46	120	200	16	180	52	117	21	62	168
203	43	90	35	37	198	133	38	136	58	192	176	215	28	171	253
60	173	43	77	169	151	148	188	134	188	76	5	211	62	207	55
165	156	127	144	210	226	82	208	186	55	45	44	114	144	234	20
44	141	63	218	151	48	210	37	50	188	78	100	66	83	120	225
202	89	201	175	183	99	58	125	171	78	232	81	9	110	238	185
21	223	53	6	66	165	35	185	41	42	81	35	66	150	201	104
68	244	63	124	221	208	186	126	236	14	230	11	184	224	209	58

34	190	74	206	29	42	171	196	57	131	13	226	53	29	140	190
16	149	250	131	103	182	200	194	3	183	181	19	62	128	177	61
107	217	242	176	61	164	124	112	177	56	234	167	60	190	102	152
2	205	77	188	160	140	243	72	13	118	184	20	27	28	216	119
150	93	173	227	45	85	4	13	109	83	190	183	254	44	116	147
247	68	119	196	192	125	251	245	202	227	175	255	240	28	233	185
137	237	225	186	187	144	82	220	85	56	15	82	136	86	86	211
200	81	131	34	167	119	252	109	57	28	145	75	189	155	130	226
176	52	184	200	182	153	199	58	219	222	95	55	46	150	123	49
254	250	36	137	218	149	92	159	150	148	194	42	139	153	169	71
12	106	183	133	195	232	237	124	244	121	153	149	15	111	250	35
126	55	101	97	218	15	252	68	43	53	199	156	13	193	191	131
197	69	175	193	105	109	150	48	217	119	165	196	200	93	198	2
80	242	122	48	126	88	249	176	21	96	189	108	223	20	103	0
212	120	170	72	142	205	146	144	218	118	24	199	36	133	143	97
3	1	138	154	44	133	195	9	167	180	153	230	18	232	230	129

$$
\begin{matrix}
96 & 49 & 188 & 112 & 107 & 141 & 222 & 157 & 170 & 205 & 46 & 109 & 178 & 253 & 165 & 222 \\
139 & 181 & 252 & 174 & 248 & 98 & 53 & 127 & 218 & 66 & 139 & 137 & 250 & 100 & 150 & 187 \\
108 & 151 & 14 & 72 & 145 & 228 & 52 & 53 & 70 & 105 & 19 & 118 & 36 & 191 & 156 & 146 \\
92 & 91 & 46 & 174 & 129 & 134 & 28 & 84 & 214 & 192 & 149 & 81 & 53 & 192 & 186 & 15 \\
154 & 238 & 238 & 40 & 35 & 232 & 177 & 185 & 167 & 104 & 28 & 48 & 208 & 240 & 93 & 15 \\
22 & 57 & 33 & 35 & 108 & 80 & 156 & 75 & 102 & 41 & 230 & 146 & 7 & 207 & 233 & 195 \\
238 & 44 & 12 & 225 & 133 & 232 & 13 & 38 & 73 & 103 & 162 & 224 & 112 & 129 & 227 & 153 \\
203 & 197 & 72 & 114 & 207 & 99 & 62 & 144 & 43 & 25 & 9 & 33 & 78 & 111 & 84 & 171 \\
163 & 174 & 140 & 226 & 76 & 105 & 49 & 52 & 55 & 55 & 78 & 78 & 120 & 67 & 2 & 121 \\
73 & 122 & 80 & 143 & 105 & 146 & 148 & 111 & 136 & 29 & 174 & 98 & 78 & 119 & 51 & 229 \\
195 & 191 & 32 & 244 & 64 & 42 & 185 & 129 & 215 & 129 & 33 & 4 & 253 & 106 & 132 & 236 \\
150 & 135 & 175 & 43 & 43 & 30 & 79 & 76 & 184 & 216 & 135 & 150 & 255 & 160 & 105 & 253 \\
216 & 116 & 114 & 9 & 20 & 109 & 72 & 238 & 216 & 14 & 215 & 228 & 172 & 248 & 98 & 27 \\
162 & 203 & 160 & 20 & 89 & 234 & 236 & 104 & 233 & 156 & 240 & 151 & 239 & 148 & 68 & 168 \\
8 & 161 & 190 & 31 & 14 & 189 & 213 & 1 & 207 & 246 & 69 & 125 & 94 & 13 & 254 & 154 \\
132 & 115 & 175 & 134 & 60 & 136 & 18 & 161 & 2 & 52 & 249 & 201 & 39 & 86 & 62 & 122
\end{matrix}
$$

$$
\begin{matrix}
175 & 213 & 230 & 188 & 248 & 27 & 35 & 68 & 34 & 106 & 240 & 15 & 74 & 205 & 3 & 192 \\
110 & 131 & 39 & 230 & 166 & 152 & 240 & 255 & 197 & 110 & 230 & 25 & 33 & 96 & 130 & 43 \\
184 & 106 & 138 & 210 & 251 & 94 & 208 & 57 & 174 & 201 & 215 & 106 & 108 & 174 & 243 & 175 \\
185 & 50 & 151 & 140 & 253 & 90 & 4 & 216 & 206 & 172 & 143 & 243 & 115 & 120 & 45 & 13 \\
251 & 101 & 66 & 108 & 54 & 90 & 42 & 250 & 69 & 147 & 82 & 244 & 7 & 252 & 179 & 53 \\
246 & 79 & 17 & 51 & 226 & 3 & 176 & 86 & 114 & 154 & 93 & 127 & 85 & 175 & 139 & 80 \\
117 & 210 & 13 & 36 & 64 & 52 & 191 & 216 & 132 & 251 & 226 & 96 & 201 & 235 & 189 & 122 \\
144 & 9 & 201 & 125 & 213 & 216 & 83 & 64 & 136 & 217 & 242 & 64 & 255 & 26 & 66 & 141 \\
214 & 245 & 158 & 201 & 168 & 139 & 68 & 3 & 221 & 20 & 135 & 142 & 208 & 182 & 145 & 192 \\
152 & 34 & 210 & 198 & 251 & 191 & 3 & 146 & 82 & 162 & 51 & 157 & 160 & 224 & 65 & 142 \\
10 & 175 & 11 & 7 & 194 & 247 & 249 & 194 & 177 & 63 & 246 & 102 & 49 & 206 & 80 & 30 \\
97 & 182 & 174 & 42 & 88 & 184 & 216 & 221 & 242 & 61 & 93 & 2 & 195 & 56 & 88 & 186 \\
121 & 190 & 103 & 125 & 218 & 102 & 182 & 84 & 59 & 20 & 67 & 116 & 220 & 245 & 157 & 187 \\
197 & 238 & 119 & 91 & 129 & 217 & 7 & 121 & 205 & 189 & 158 & 210 & 44 & 189 & 62 & 69 \\
208 & 216 & 180 & 176 & 14 & 27 & 146 & 157 & 214 & 11 & 150 & 20 & 19 & 162 & 208 & 139 \\
47 & 248 & 48 & 34 & 135 & 186 & 60 & 178 & 108 & 255 & 230 & 254 & 58 & 65 & 30 & 66
\end{matrix} \tag{5.2}
$$

$$
\begin{bmatrix}
113 & 73 & 66 & 92 & 33 & 16 & 91 & 0 & 52 & 245 & 249 & 45 & 45 & 131 & 17 & 48 \\
163 & 158 & 75 & 34 & 247 & 172 & 222 & 169 & 121 & 200 & 217 & 190 & 113 & 118 & 23 & 136 \\
98 & 91 & 235 & 68 & 203 & 52 & 99 & 66 & 36 & 60 & 125 & 77 & 109 & 157 & 33 & 14 \\
101 & 252 & 70 & 162 & 63 & 209 & 94 & 80 & 78 & 75 & 208 & 1 & 119 & 112 & 66 & 3 \\
115 & 55 & 85 & 16 & 102 & 144 & 138 & 114 & 254 & 13 & 61 & 230 & 165 & 215 & 168 & 126 \\
149 & 113 & 194 & 100 & 34 & 60 & 85 & 86 & 117 & 204 & 242 & 107 & 29 & 166 & 100 & 208 \\
247 & 69 & 167 & 204 & 194 & 215 & 235 & 46 & 240 & 52 & 46 & 161 & 53 & 216 & 147 & 195 \\
75 & 223 & 70 & 220 & 1 & 123 & 188 & 9 & 122 & 130 & 106 & 217 & 74 & 225 & 145 & 148 \\
188 & 77 & 47 & 145 & 165 & 250 & 126 & 42 & 175 & 39 & 141 & 45 & 186 & 11 & 78 & 122 \\
124 & 108 & 85 & 97 & 134 & 37 & 232 & 80 & 170 & 252 & 236 & 134 & 228 & 6 & 15 & 229 \\
106 & 242 & 28 & 236 & 187 & 64 & 255 & 132 & 233 & 145 & 78 & 54 & 237 & 17 & 214 & 126 \\
105 & 184 & 24 & 1 & 163 & 238 & 34 & 79 & 142 & 213 & 185 & 81 & 233 & 98 & 6 & 91 \\
109 & 12 & 148 & 237 & 225 & 180 & 125 & 20 & 254 & 196 & 192 & 104 & 21 & 54 & 125 & 40 \\
33 & 15 & 59 & 207 & 172 & 241 & 219 & 196 & 156 & 214 & 230 & 250 & 71 & 163 & 9 & 229 \\
3 & 95 & 140 & 134 & 160 & 30 & 140 & 95 & 94 & 174 & 151 & 224 & 47 & 87 & 52 & 233 \\
34 & 38 & 184 & 252 & 222 & 57 & 78 & 47 & 46 & 3 & 30 & 96 & 108 & 156 & 203 & 26
\end{bmatrix} \tag{5.3}
$$

Data Compression for Video-Conferencing using Half tone and Wavelet Transform

Dr. H.B.Kekre
Sr. Professor, Computer Engineering, NMIMS University,
Mumbai-400056, India

Sanjay R. Sange
Assistant Professor, Information Technology, NMIMS University,
Mumbai-400056, India

Dr. Tanuja K. Sarode
Associate Professor, Computer Engineering,
TSEC, Mumbai University
Mumbai, India

***Abstract*—Overhead of data transmission over internet is increasing exponentially every day. Optimization of natural bandwidth is the basic motive by compressing image data to the maximum extend. For the same objective, combination of lossy half tone and lossless Wavelet Transform techniques is proposed so as to obtain low-bit rate video data transmission. Decimal values of bitmapped image are to be converted into either 1 or 0 in half toning process that incur pictorial loss and gives 8:1 compression ratio (CR) irrespective of image. Wavelet Transform is applied on half tone image for higher compression for various levels. An experimental result shows the higher CR, minimum Mean Square Error (MSE). Ten sample images of different people captured by Nikon camera are used for experimentation. All images are bitmap (.BMP) 512 X 512 in size. The proposed technique can be used for video conferencing, storage of movies and CCTV footage etc.**

Keywords-Half tone; Low-Bit rate; video data compression; Wavelet Tranform; Bandwidth optimization; Structural Similarity Index Measure (SSIM).

I. INTRODUCTION

From last two decades Wavelet Transform has found enormous application in different areas like speech, computer graphics, signal, image processing and in medical field for DNA, ECG, protein, blood pressure, and heart rate analysis. Wavelet Transform overcomes the limitations of Fourier Transform as it cannot detect local properties as in [1]. Hybrid Wavelet Transform using any two orthogonal transform can be used for higher image data compression with minimum loss using a set of complimentary wavelets, where comparison of DCT, DHT, DWT and Kekre transform is explained as in [2]. The combination of Wavelet Transform with Modified-Run-Length- Coding (MRLC) along with new quantization technique is proposed for ECG data compression. This proposed method improves data compression by 13 % as in [3].

Generation of Wavelet Transform from any orthogonal transforms by contraction and translation infinite set of functions can be generated. Experimental results of original image with reconstructed image using orthogonal transforms Walsh and DCT with respect to their Wavelets are compared. Walsh Wavelet and DCT Wavelet results are better than Walsh and DCT as in [4].

Wavelet Transform for high resolution satellite imageries with lifting scheme is proposed that reduces computational time and resources with appreciable results as in [5]. Considering main three factors of high embedding capacity, imperceptibility and robustness effective stenography is explained with Walsh Wavelet and DCT Wavelet proven that are prone to filtering, noise, cropping and compression of an image as in [6]. Spikes at different frequencies and amplitude using Wavelet Transform for Neural data compression from different channels are found to reconstruct unique signature and relate it some activities as in [7]. Various orthogonal Wavelet transforms of Walsh, Cosine, Hartley, Kekre are used for image data compression and proved better results as compared to respective normal forms. 70% to 90% is compressed by removing low energy coefficients in their respective Wavelet forms as in [8].

Section II explains about the half tone method and various half tone operators, Section III explains the Hybrid Algorithm of Half tone and Wavelet Transform, Section IV explains about experimental results and discussion. In section V paper conclusion and future scope is explained.

II. HALF TONING METHOD

A. Neighbourhood Processing

Half toning is the process in which intensity and pattern of dot varies to simulate different shades. Half tone dots are produced by superimposing mask over the image. Half toning is the error diffusion process that results into noisy image. Half toning templates shown in fig.1a to fig.1d are used to convert continuous tone image into half tone image. These templates are rotated on continues tone image as neighborhood processing.

For the same objective of high image data compression and low-bit-rate data transmission to optimize bandwidth for video conferencing, other techniques are used as hybrid technique with half tone technique. Half toning is the lossy technique and gives 8:1 CR. Two-fold hybrid techniques are used for higher compression ratio with half tone. Half tone with Kekre's Fast Codebook Generation (KFCG) vector quantization technique is presented by Kekre et al as in [9]. Lossy half tone with lossless Huffman coding technique is presented as in [10]. Lossy half tone with lossless Run-Length-Encoding technique is presented as in [11]. Importance of red plane from time complexity point of view is explained as in [12]. For reconstruction of image from half tone image Inverse half toning algorithm as in [13] is described. Some other half toning operators are proposed with performance analysis as in [14].

0	0	0
0	X	7
3	5	1

Fig. 1a: Floyd-Steinberg

0	0	0
0	X	1
0	1	3

Fig. 1b. Small

0	0	0	0	0
0	0	0	0	0
0	0	X	7	5
3	5	7	5	3
1	3	5	3	1

Fig. 1c.: Jarvis

0	0	0	0	0
0	0	0	0	0
0	0	X	1	9
23	7	5	3	11
21	19	17	15	13

Fig. 1d.: South-East

Figure 1. Half tone operators: Fig.1a is Floyd-Steinberg half tone operator, Fig.1b is Small half tone operator, Fig.1c is Jarvis half tone operator, Fig.1d is South-East half tone operator

B. *Quantization*

As shown fig.2 color image is split into three primary R-G-B planes and it posses gray levels from 0-255, representing each pixel by 8-bit. After half tone technique, quantization process is used to convert gray level into bi-level with loss as either 0 or 1 as in [13].

III. Hybrid Half Tone With Wavelet Algorithm

As shown in fig.2, on each plane of half tone image Haar Wavelet transform is applied. Fig.3 shows the working principle of Wavelet transform.

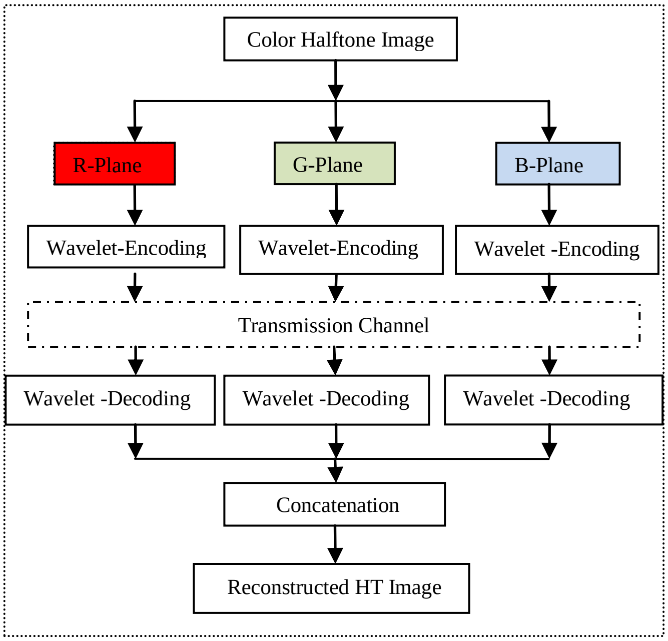


Figure 2. Block diagram of Half tone-Wavelet Transform

A. *Algorithm*

- Wavelet transform is applied on binary half tone image of size 512-by-512.
- In Wavelet transform alternate row and column is eliminated so as to reduce the overall NXN size of an image as is represented in fig.3 and sample result is shown in fig.4. b

$$Li=N/2^{i} \qquad (1)$$

Where i =1, 2 … n

- Wavelet transform is applied on half tone image plane till to the desired level of compression. At each level of compression Wavelet transform extract features at different frequencies and location.
- Wavelet transform encoded data in its highest compressed form can be used for transmission on channel.
- At the receiving end inverse Wavelet transform is applied to decode image data so as to obtain half tone image.
- Inverse half toning algorithm is applied with concatenation of all the half tone planes to reconstruct of an image.

B. *Compression Ratio (CR)*

In first iteration of Wavelet transform image size of 512-by-512 in half tone form is converted into 256-by-256 as level-1 and referred as L1. At L1 Wavelet transform compresses data 50% to that of half tone image data as shown in fig. 3. Eq. (1) shows the decomposition of NXN size image into desired level Li. The Wavelet Transform is applied on low-low frequency component of L1, for further compression of 50% data is called as L2 and image size becomes 128-by-128 and henceforth L3 and L4 is achieved. The CR of original image to inverse image at L1 is 32. In the same way, L2 compresses original image data 128 times, L3 compresses 512 times and L4 compresses 2048 times.

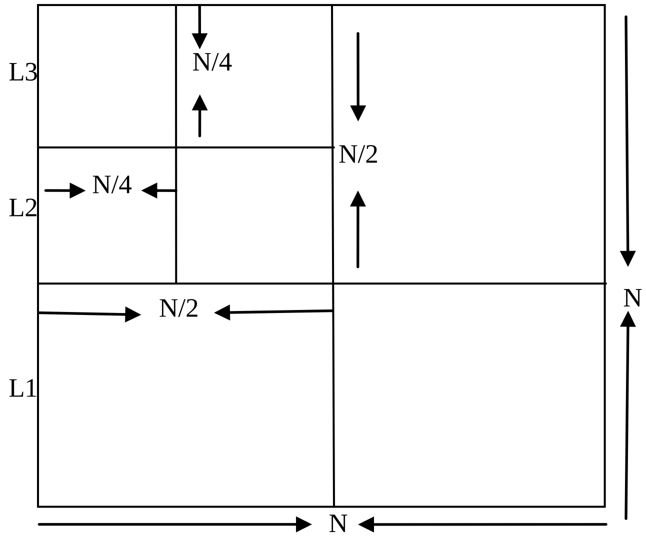


Figure 3. Wavelet Transform Pyramid

IV. Results And Discussion

Combination of Half tone with Wavelet Transform is used to achieve higher CR on different ten 512-by-512 in size bitmapped images, for low bit-rate image data compression in Video-Conferencing. Mean-Square-Error (MSE) and Structural Similarity Index Measure (SSIM) are used between original image and inverse image.

Table-I, III, V and Table-VII shows the MSE between original images and inverse image for different half tone operators at L1, L2, L3 and L4. As well as Table-II, IV, VI and Table-VIII shows the SSIM between original image and inverse images for different half tone operators at L1, L2, L3 and L4. Fig. 5, 7, 9 and 11 shows the graphical representation of MSE between original images and inverse image for different half tone operators at L1, L2, L3 and L4. As well as fig. 6, 8, 10 and 12 shows the SSIM between original image and inverse images for different half tone operators at L1, L2, L3 and L4.

Fig. 13 shows the reconstructed images from different half tone operators and Wavelet Transform at levels from L1 to L4. Reconstructed image quality of Small operator is almost same to that of standard Floyd-Steinberg and Jarvis operators. As well as it gives same MSE and SSIM with reduced computational complexity [11].

Whereas South-East operator gives higher MSE and negligible poor in image quality as compared to standard operators. Fig. 13 shows the MSE increases and image quality decreases from L1 to L4. As shown in fig.3, in each level only Low-Low frequency component is used to take inverse Wavelet transform. Remaining Low-High, High-Low and High-High components are considered as matrix of zeros of the same size in the respective level.

TABLE I. MSE- Between Inverse And Original Image Using Different Half Tone Operators Wavelet Transform At-L1

S.N.	Image	Flyod	Jarvis	Small	South-East
1	Aditi	117.8627	215.8853	249.986	360.656
2	KekreHB	318.1064	264.8021	391.9745	447.651
3	Sanjay	285.8108	377.4874	398.578	621.227
4	Anita	263.2586	452.0141	506.7903	903.042
5	Tandle	142.0714	342.3254	389.9147	777.652
6	Pallavi	130.8618	187.4659	171.0308	633.804
7	More	172.1179	297.9908	316.0038	625.25
8	Shruti	108.5584	177.0217	173.92	646.386
9	Ravi	302.4952	224.187	287.0564	625.001
10	Ajay	214.7108	195.7814	275.4807	577.054
	Average	205.5854	273.49611	316.07352	621.772

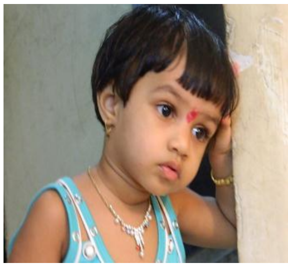


Figure 4. (a).Original image: Aditi

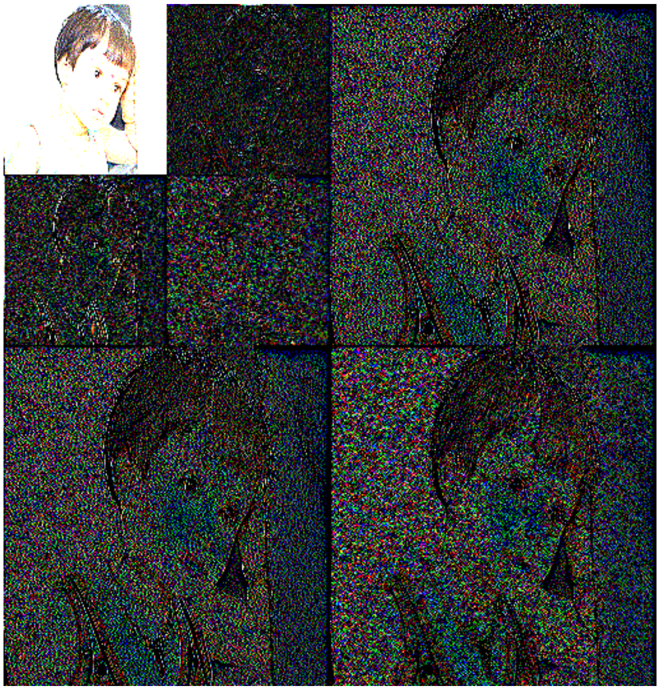


Figure 4. (b). Aditi: Wavelet Transform (L3) image using Jarvis half tone

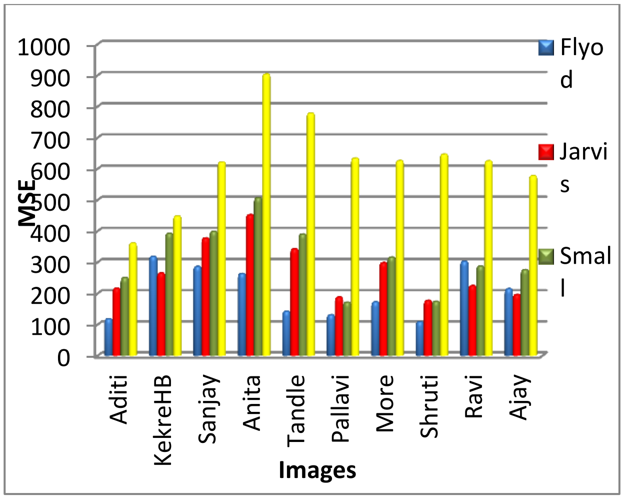


Figure 5. MSE between inverse and original images using different half tone operator and Wavelet Transorm at L1

TABLE II. SSIM- BETWEEN INVERSE AND ORIGINAL IMAGE USING DIFFERENT HALF TONE OPERATORS WAVELET TRANSFORM AT-L1

S.N	Image	Flyod	Jarvis	Small	South-East
1	Aditi	0.9911	0.9906	0.9911	0.9936
2	KekreHB	0.966	0.9931	0.9798	0.999
3	Sanjay	0.9957	0.9964	0.9968	0.9982
4	Anita	0.9893	0.9909	0.9909	0.9939
5	Tandle	0.9841	0.983	0.9842	0.9872
6	Pallavi	0.998	0.9995	0.9996	1
7	More	0.9936	0.9929	0.9936	0.9955
8	Shruti	0.9958	0.9975	0.9978	0.9996
9	Ravi	0.9821	0.9975	0.9989	0.9992
10	Ajay	0.9769	0.9948	0.9946	0.9969
	Average	0.98726	0.99362	0.99273	0.99631

TABLE III. MSE- BETWEEN INVERSE AND ORIGINAL IMAGE USING DIFFERENT HALF TONE OPERATORS WAVELET TRANSFORM AT-L2

S.N.	Image	Flyod	Jarvis	Small	South-East
1	Aditi	117.9284	215.8858	249.9898	360.688
2	KekreHB	318.1064	264.8021	391.9745	447.651
3	Sanjay	285.8108	377.4875	398.578	621.227
4	Anita	263.2586	452.0141	506.7903	903.045
5	Tandle	142.0714	342.3254	389.9147	777.652
6	Pallavi	130.8618	187.4659	171.0308	633.804
7	More	172.1183	297.991	316.0038	625.25
8	Shruti	108.5584	177.0218	173.9202	646.386
9	Ravi	302.4952	224.187	287.0564	625.001
10	Ajay	214.7103	195.7814	275.4808	577.054
	Average	205.592	273.4962	316.07393	621.776

TABLE IV. SSIM- BETWEEN INVERSE AND ORIGINAL IMAGE USING DIFFERENT HALF TONE OPERATORS WAVELET TRANSFORM AT-L2

S.N.	Image	**Flyod**	**Jarvis**	**Small**	**South-East**
1	Aditi	0.9911	0.9906	0.9911	0.9937
2	KekreHB	0.966	0.9931	0.9798	0.999
3	Sanjay	0.9957	0.9964	0.9968	0.9982
4	Anita	0.9893	0.9909	0.9909	0.9939
5	Tandle	0.9841	0.983	0.9842	0.9872
6	Pallavi	0.998	0.9995	0.9996	1
7	More	0.9936	0.9929	0.9936	0.9955
8	Shruti	0.9958	0.9975	0.9978	0.9996
9	Ravi	0.9821	0.9975	0.9989	0.9992
10	Ajay	0.9769	0.9948	0.9946	0.9969
	Average	0.98726	0.99362	0.99273	0.99632

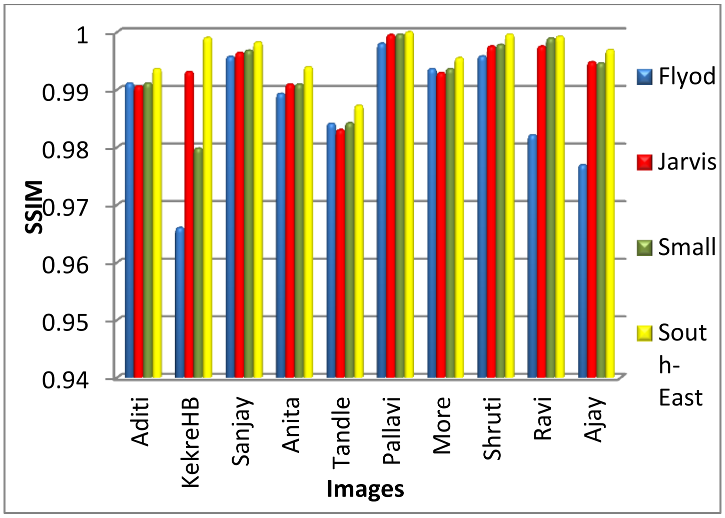


Figure 6. SSIM between inverse and original images using different half tone operator and Wavelet Transom at L1

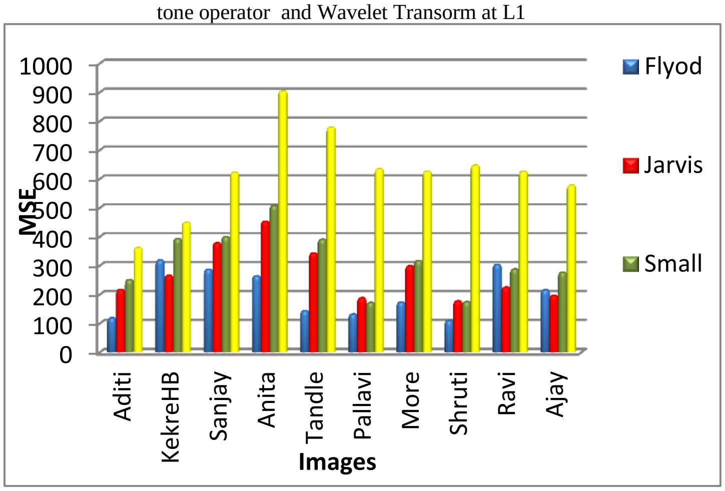


Figure 7. MSE between inverse and original images using different half tone operator and Wavelet Transorm at L2

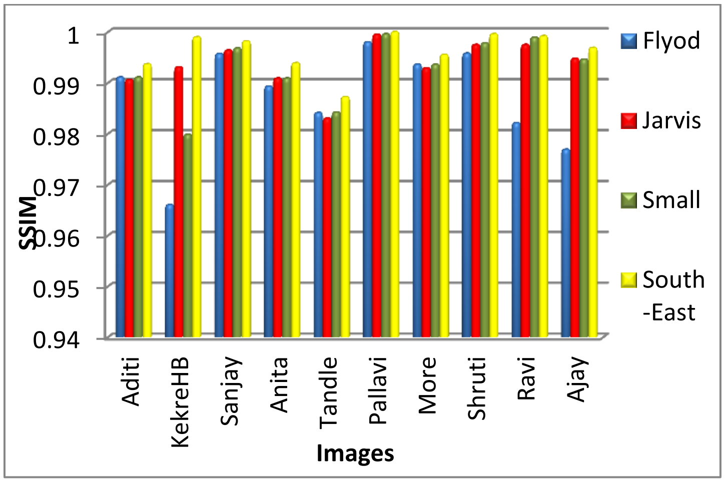


Figure 8. SSIM between inverse and original images using different half tone operator and Wavelet Transorm at L2

TABLE V. MSE- BETWEEN INVERSE AND ORIGINAL IMAGE USING DIFFERENT HALF TONE OPERATORS WAVELET TRANSFORM AT-L3

S.N.	Image	Flyod	Jarvis	Small	South-East
1	Aditi	134.267	220.2996	227.7553	370.799
2	KekreHB	368.0072	224.9916	403.7703	435.893
3	Sanjay	312.8477	345.6988	330.1543	461.705
4	Anita	203.2026	264.9968	302.8949	542.179
5	Tandle	121.0674	228.9125	226.9721	412.069
6	Pallavi	163.8037	187.3523	169.3533	381.113
7	More	171.3309	257.8598	236.361	428.801
8	Shruti	155.6485	194.6737	175.9346	375.641
9	Ravi	194.6619	190.1707	207.5549	436.251
10	Ajay	121.9036	184.7462	195.4543	420.552
	Average	194.6741	229.9702	247.6205	426.5

TABLE VI. SSIM- BETWEEN INVERSE AND ORIGINAL IMAGE USING DIFFERENT HALF TONE OPERATORS WAVELET TRANSFORM AT-L3

S.N.	Image	**Flyod**	**Jarvis**	**Small**	**South-East**
1	Aditi		0.9905	0.991	0.9935
2	KekreHB	0.9583	0.9998	0.972	1
3	Sanjay	0.9927	0.9964	0.9968	0.9982
4	Anita	0.9873	0.9908	0.9868	0.9938
5	Tandle	0.984	0.9831	0.9841	0.9873
6	Pallavi	0.9968	0.9995	0.9996	1
7	More	0.9935	0.9929	0.9935	0.9955
8	Shruti	0.9909	0.9974	0.9978	0.9989
9	Ravi	0.9879	0.9963	0.999	0.9995
10	Ajay	0.988	0.9947	0.9917	0.9969
	Average	0.9866	0.99414	0.99123	0.9964

Image data can be compressed to the higher level of Wavelet Transform say 16X16 and 8X8, but the image reconstruction quality degrades a lot.

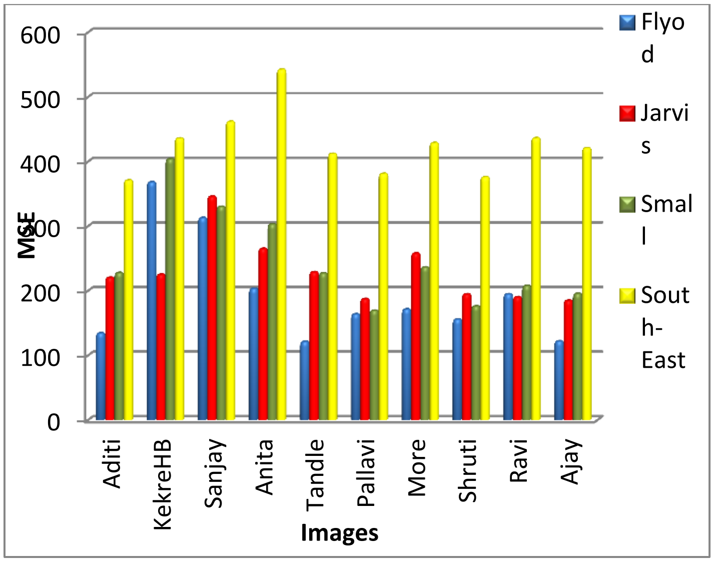


Figure 9. MSE between inverse and original images using different half tone operator and Wavelet Transorm at L3

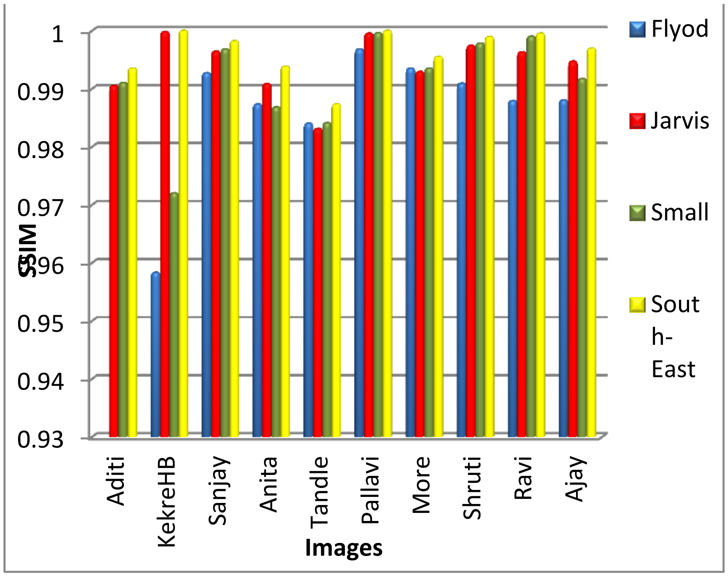


Figure 10. SSIM between inverse and original images using different half tone operator and Wavelet Transorm at L3

TABLE VII. MSE- Between Inverse And Original Image Using Different Half Tone Operators Wavelet Transform At-L4

		Flyod	Jarvis	Small	South-East
1	Aditi	190.86	240.092	225.3853	345.85
2	KekreHB	595.84	510.291	586.0695	360.28
3	Sanjay	703.25	458.67	489.1887	541.14
4	Anita	380	254.844	259.2035	378.81
5	Tandle	231.56	204.578	197.7312	308.28
6	Pallavi	232.66	199.918	184.2595	310.61
7	More	275.28	243.013	241.0863	342.42
8	Shruti	231.27	203.948	186.0357	300.9
9	Ravi	978.47	235.499	480.6849	325.47
10	Ajay	525.64	178.201	252.2813	311.26
	Average	434.48	272.905	310.1926	352.5

TABLE VIII. SSIM- Between Inverse And Original Image Using Different Half Tone Operators Wavelet Transform At-L4

		Flyod	Jarvis	Small	South-East
1	Aditi	0.991	0.9903	0.9908	0.9933
2	KekreHB	0.9409	0.9553	0.9487	0.9842
3	Sanjay	0.992	0.9962	0.9966	0.9981
4	Anita	0.9759	0.9907	0.9847	0.9938
5	Tandle	0.984	0.9831	0.984	0.9873
6	Pallavi	0.9954	0.9994	0.9996	1
7	More	0.9936	0.9929	0.9935	0.9955
8	Shruti	0.9877	0.9974	0.9978	0.9989
9	Ravi	0.9573	0.9941	0.9759	1
10	Ajay	0.9605	0.9947	0.9825	0.9977
	Average	0.97783	0.98941	0.98541	0.99488

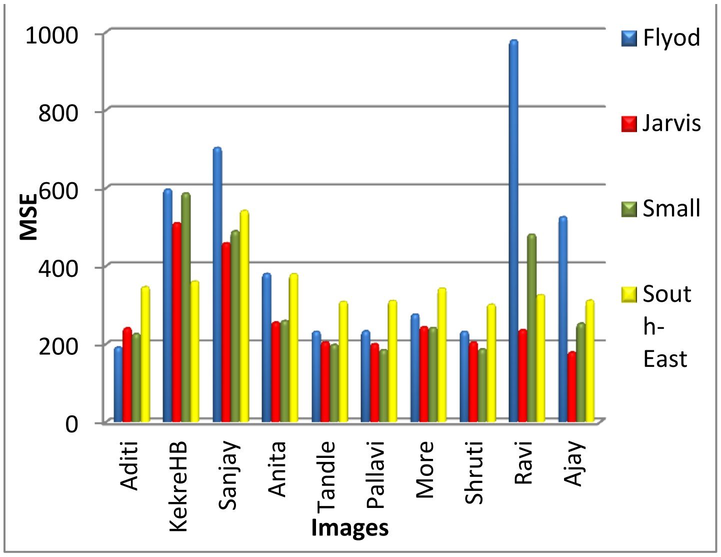


Figure 11. MSE between inverse and original images using different half tone operator and Wavelet Transorm at L4

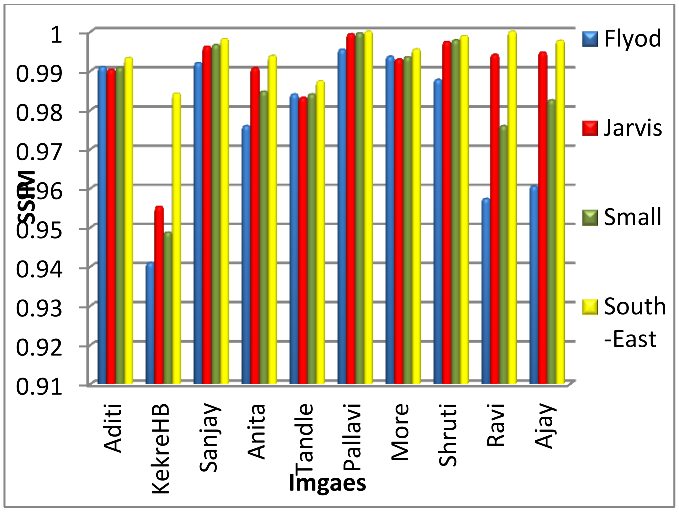


Figure 12. SSIM between inverse and original images using different half tone operator and Wavelet Transorm at L4

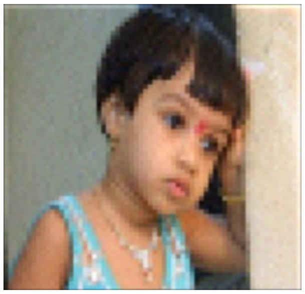
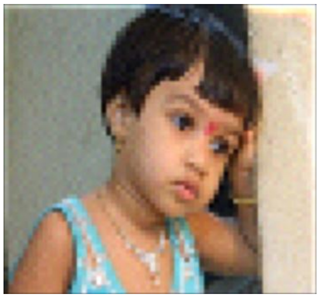
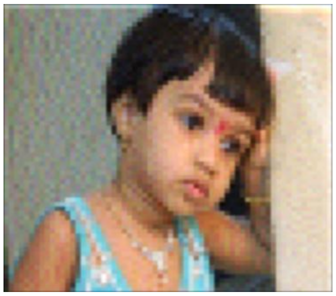
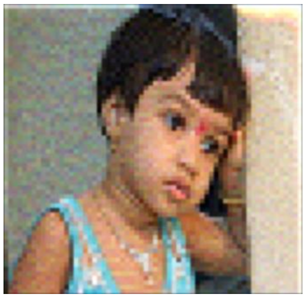
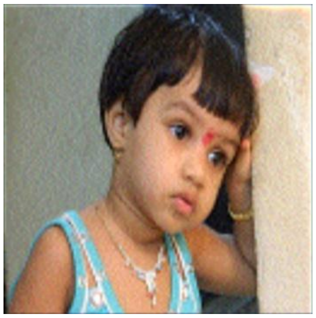
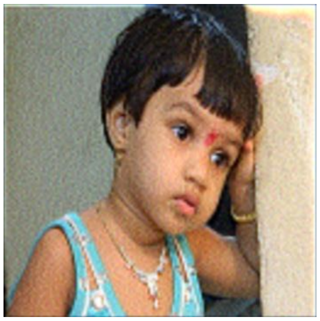
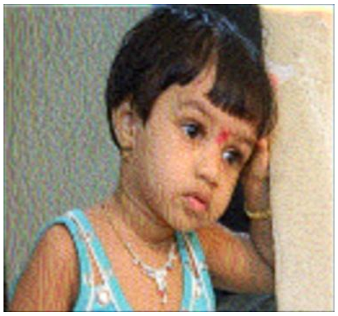
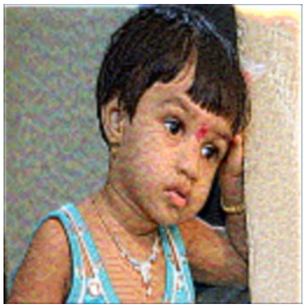
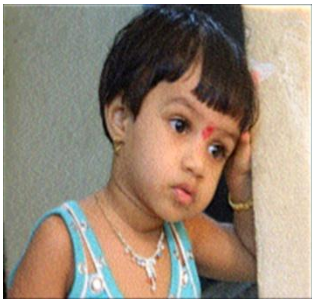
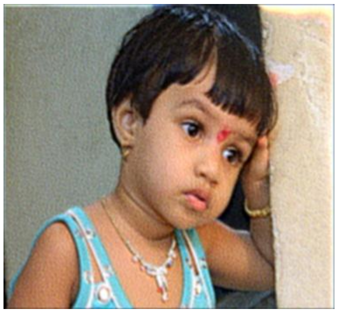
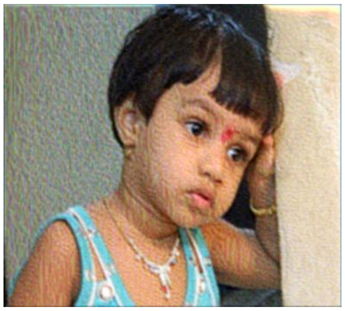
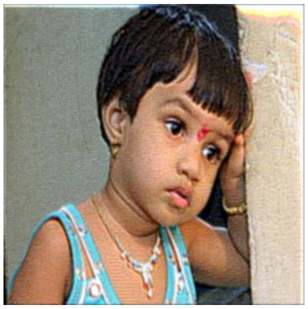
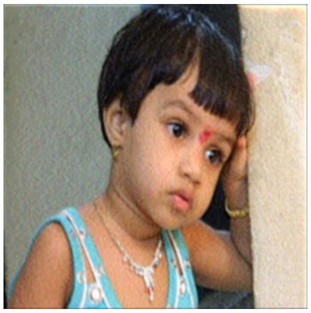
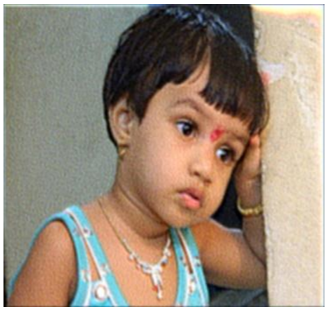
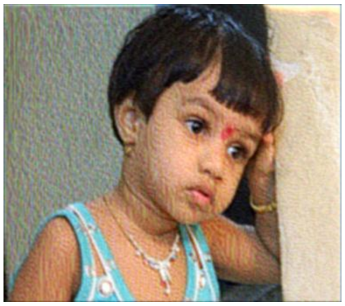
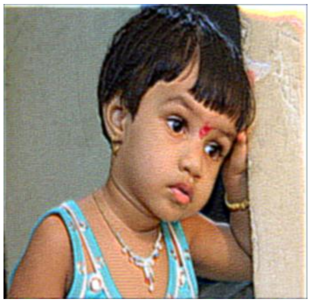
Size	Floyd-Steinberg	Jarvis	Small	South-East
32X32 L4	 a. MSE=190.8560	 b. MSE=240.0915	 c. MSE=225.3853	 d. MSE= 345.8487
64X64 L3	 e. MSE=134.2670	 f. MSE=220.2996	 g. MSE=227.7553	 h. MSE=370.7987
128X128 L2	 i. MSE=117.9284	 j. MSE=215.8858	 k. MSE=249.9898	 l. MSE=360.688
256X256 L1	 m. MSE=117.8627	 n. MSE= 215.8853	 o. MSE=249.986	 p. MSE=360.656

Figure 13. Reconstructed images using different half tone operators and Wavelet Transform at different levels from L1 to L4:

(1) CR=32 for L1, (2) CR=128 for L2, (3) CR=512 for L3, (4) CR=2048 for L4

V. CONCLUSION

For low-bit rate video data transmission image data is compressed using combination of half tone and Wavelet Transform on ten different 512 by 512 bitmap images. Wavelet Transform is applied at different levels that converts image from 512 by 512 to 256 by 256 as level L1, 128 by 128 as L2, 64 by 64 as L3 and 32 by 32 as L4.

Below this level reconstructed image quality degrades. Future scope to this paper is to convert Wavelet Transform domain image into desired number of non-overlapping blocks. Calculate energy of all the blocks and can eliminate the some lowest energy blocks based on threshold. Elimination of such non-overlapping blocks will increase the CR. In real-time processing, proposed algorithm takes more processing time as compared to the frame rate that required for smooth video-conferencing. As well as to develop an algorithm to preserve the features of Low-High, High-Low and High-High components of Wavelet transform domain image and can be added to Low-Low frequency component of Wavelet transform.

REFERENCES

[1] S.Mallat,"ATheory of Multiresolution Signal Decomposition: The Wavelet Representation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.11,pp.674-693, 1989

[2] H.B.Kekre, T.K. Sarode Sudeep Thepade,"Inception of Hybrid Wavelet Tranform using Two Orthogonal Transforms and It's use for Image Compression", IJCSIS, vol.9, issue-6, pp.80–87.

[3] Hsieh-Wei Lee, King-Chu Hung, Tsung-Ching Wu, ChengTung Ku," A Modified Run-length Coding for the realization of Wavelet-based ECG Data Compression System", Journal of Next Generation Information Technology, Vol. 1. Number 1, May2010

[4] H.B.Kekre, Archna Athwale, Dipali Sadavarti, "Algorithm to Generate Wavelet Transform from an Orthogonal Transform", International Journal Of Image Processing (IJIP)", vol. 4, issue 4

[5] K Nagmani and AG Anath,"Image Compression Techniques for High Resolution Satellite Imageries using Classical Lifting Scheme", International Journal of Computer Applications(0975-8887), volume 15-No.13, February 2011

[6] H.B.Kekre, Archna Patankar and Dipali Koshti." Performance Comparision of simple Orthogonal Transform and Wavelet Tranforms for Image Stegnography", Inernation Journal of Computer Applications(0975-8887), volume 44-No.6, February 2012

[7] Seetharam Narasimhan, Massod Tabib," Neural Data Compression witn Wavelet Transform: A Vocabalry Based Approch", Procedings of 3rd International IEEE EMBS Conferenceon Neural engineering Kohal Coast, Hawaii USA, May2-5, 2007

[8] H.B.Kekre, T.K. Sarode ,Sudeep Thepade, Sonal Shroff,"Instigation of Orthogonal Wavelet Transforms using Walsh, Cosine, Hartley, Kekre Transforms and their use in Image Compression", (IJCSIS) International Journal of Computer Science and Information Security, Vol.XXX, No.XXX,2010

[9] H.B. Kekre, Tanuja K. Sarode, Sanjay R. Sange, Shachi Natu, and Prachi Natu , "Halftone Image Data Compression using Kekre's Fast Code Book Generation (KFCG) Algorithm for Vector Quantization", International Conference, ICTSM 2011, CCIS 145, pp. 34–42, 2011 © Springer-Verlag Berlin Heidelberg 2011, ICTSM 2011

[10] H. B. Kekre, Sanjay R. Sange, Gauri S. Sawant, and Ankit A. Lahoty International Conference, "Image Data Compression using Halftone and Huffman Coding", ICTSM 2011, CCIS 145, pp. 221–226, 2011 © Springer-Verlag Berlin Heidelberg 2011

[11] H.B. Kekre, Tanuja K. Sarode, Sanjay R. Sange, Pallavi Halankar,"New Half tone Operators for High Data Compression in Video- Conferencing", International Conference ICSCA 2012, on 9th and 10th June 2012, Singapore, published in International Proceedings of Computer Science and Information Technology (IPCSIT), volume 41, pg. 211-218, available online http://www.ipcsit.com/vol41/038-ICSCA2012-S3001.pdf

[12] H.B. Kekre, Tanuja K. Sarode, Sanjay R. Sange, Bhumika Raghwani , Merlin Vergis "Halftone Image Data Compression using KFCG Vector Quantization Algorithm for Video Conferencing", Journal of Signal and Image Processing, ISSN: 0976-8882 and E-ISSN: 0976-8890, Vol-2, Issue2, 2011, pp-42-49. Available online at http://www.bioinfo.in/contents.php?id=48

[13] H.B.Kekre, Sanjay R.Sange, "Restoration of Color Halftone image by using Fast Inverse Half toning Algorithm" International Conference on "Advances in Recent Technologies in Communication and Computing, (ARTCom 2009)" to be held on 27-28 October 2009 at Kottayam, Kerla, organized by ACEEE, Los Alamitos, California (U.S). 978-0-7695-3845-7/09

[14] H.B.Kekre, M.U.Kharat, Sanjay R. Sange , "Image data compression using new Halftoning operators and Run Length Encoding", International Conference on "Contours of Computing Technology, (THINKQUEST 2010)",pg-224 to pg-230 to be held on 13th-14th March 2010

AUTHORS PROFILE

Dr. H. B. Kekre B.E. (Hons.) in Telecomm. Engg. in 1958, M.Tech from IIT Bombay in 1960, M.S. Engg. from University of Ottawa in 1965 and Ph.D. from IIT Bombay in 1970. He is currently Senior Professor working with MPSTME, NMIMS, Mumbai INDIA. He has guided 17 Ph.D.s, 150 M.E./M.Tech Projects and several B.E./B.Tech Projects. His areas of interest are Digital Signal processing, Image Processing and Computer Networks. He has more than 400 papers in National/International Conferences/Journals to his credit. Recently his 12 students have received best paper awards. Seven of his students have been awarded Ph. D. of NMIMS. Currently he is guiding ten Ph.D. students. He is member of ISTE and IETE.

Dr. Tanuja K. Sarode has received M.E.(Computer Engineering) from Mumbai University in 2004, Ph.D. from MPSTME SVKM's NMIMS, Mumbai, INDIA. She has more than 11 years of experience in teaching. She is currently working as Assistant Professor in Dept. of Computer Engineering at TSEC, Mumbai. She is member of IAENG and IACSIT. Her areas of interest are Image Processing, Signal Processing and Computer Graphics. She has more than 100 papers in National/International Conferences/journal to her credit.

Sanjay Ramkrishna Sange has received M.E. (Computer Engineering) from Mumbai University in 2008.He is the member of ISTE. He is pursuing his PhD. from NMIMS, Mumbai. He has presented/published 15 papers in National, International Conference/Journal. For one paper he has got "Best Paper" award in 2010.

Finding Association Rules through Efficient Knowledge Management Technique

Anwar M. A.
College of Engineering and Computing
Al Ghurair University
Dubai Academic City, UAE

***Abstract*— One of the recent research topics in databases is Data Mining, to find, extract and mine the useful information from databases. In case of updating transactions in the database the already discovered knowledge may become invalid. So we need efficient knowledge management techniques for finding the updated knowledge from the database. There have been lot of research in data mining, but Knowledge Management in databases is not studied much. One of the data mining techniques is to find association rules from databases. But most of association rule algorithms find association rules from transactional databases. Our research is a further step of the Tree Based Association Rule Mining (TBAR) algorithm, used in relational databases for finding the association rules .In our approach of updating the already discovered knowledge; the proposed algorithm Association Rule Update (ARU), updates the already discovered association rules found through the TBAR algorithm. Our algorithm will be able to find incremental association rules from relational databases and efficiently manage the previously found knowledge.**

Keywords- Data Mining; Co-occurrences; Incremental association rules; Dynamic Databases.

I. Introduction

At the very abstract level of data mining, it is part of Artificial Intelligence. One of the data mining techniques for finding useful information from the database is association rule. Association rules find the co-occurrences among item sets in the database. For example in a customer transaction database we want to find that whenever customer purchases item A, item B is purchased how many times. These co-occurrences are found through finding the large item sets. As mentioned in [1] to find the large item sets, it should be greater than the minimum support threshold, which is the minimum number of transactions from the database having that item set.

There are two issues related to association rules.

- Finding the preprocessing algorithm for association rules
- Update algorithm for association rules. The update algorithm enables to efficiently update the already discovered information .So the update algorithm depends very much on the preprocessing algorithm used.

Most of the association rules algorithms like Apriori [2], DHP [5], OCD [9] and [12] find association rules from transactional databases. In case of association rules from relational databases TBAR [10] algorithm was developed as a loosely couple approach.

The most recent algorithms for the update algorithms like FUP [3], MLUP [4], FUP2 [8], UWEP [7], and SWF [11] etc find updated association rules from the transactional databases. In our research we have developed a new update algorithm for finding the updated information from the relational database on the basis of the TBAR algorithm. Our performance study shows that the proposed solution is 2.1 to 2.3 times faster as compared to TBAR algorithm. We present an efficient algorithm, ARU, for finding association rules and apply a new knowledge management technique, to reuse the previously discovered knowledge from the relational databases. Precisely rather than finding large item sets from scratch, the large item sets found through the TBAR algorithm are stored and reused.

In association rules we find the co-occurrences among item sets through finding the large item sets. An item set is large if it is above the minimum support threshold .For example in a database if the minimum support threshold is 5%, then all the item sets from the database having more than 5% occurrence will be included in large item sets. So the main problem in maintenance of association rules is updating the large item sets. In our prototype system we have been able to update the large item sets more efficiently as compared to the previous approach of TBAR.

II. Prelimanaries

Let I = $\{i_1, i_2, \ldots\ldots, i_m\}$ be a set of literals, called items. Let D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. Each transaction is associated by an identifier, called TID. Let X be a set of items. A transaction T is said to contain X if and only if $X \subseteq T$. An association rule is an implication of the form $x \Rightarrow y$, where $x \subseteq I$, $y \subseteq I$ and $X \cap Y = \emptyset$. The rule $x \Rightarrow y$ holds in the transaction set D with confidence c if c% of transactions in D that contain x also contain y.

The rule $x \Rightarrow y$ has support s in the transaction set D if s% of the transactions in D contains $X \cup Y$. For a given pair of confidence and support threshold, the problem of mining association rules is to find out all the association rules that have confidence and support greater than the corresponding thresholds. As there is lot of research for finding the association rules, given large item sets, our focus will be to find the large item sets from the updated database. The notion of item must be redefined in a relational database. An item will be a pair a: v where a is the attribute and v is the value of a. a fundamental property of an item in a relational database is that they cannot contain more than one item per table column if a1:v1 and a2:v2 belong to an item set, then a1≠a2 which is the consequence of the First Normal Form (1NF) in databases: a relation is in 1NF if it's attribute domain contain atomic values only. This justifies our distinction between items in transactional and items in relational databases.

III. System Overview

Our algorithm is based on the TBAR algorithm, which finds the association rules from the relational database. Our incremental association rule algorithm is an improvement of that algorithm to find incremental association rules from the relational databases. We apply a new Knowledge Management technique, to find the incremental association rules from dynamic databases more efficiently as compared to finding the association rules from the database.

As shown in Figure 1, our algorithm is implemented as the data integration module to efficiently update the association rules. The large 1-item sets found through the TBAR algorithm is saved in the knowledge base .In our algorithm of update we have reused those large 1-item sets from the knowledge base and thus saved the CPU time and one scan of the database. As depicted in [6] we can couple association rule algorithm with the relational database in a number of ways. In our case we opted for the loosely coupled approach, as our data mining application process space is outside the database process space.

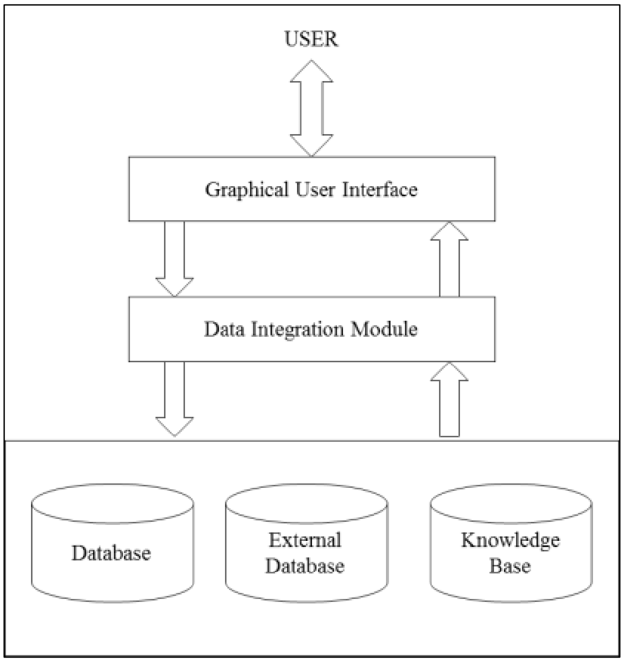


Figure 1. The System.

IV. TBAR Algorithm

The TBAR algorithm uses the item set tree data structure to efficiently store all L_{ks} .All L_{ks} are organized on the basis of levels.

TBAR Algorithm

```
Set.Init(minsup);
Itemsets=set.Relevants(1);
StoreL1(itemsets);                    (Step 4.3)
K=2;
While(k<=cols && itemsets >=k)
{
     itemsets =set.candidates(K) ;
     If(itemsets >0)
     Itemsets=set.Relevants(k);
     K++;
}
```

In this case *init* method creates and initializes the item set tree. The *set.Rrelevants(1)* method finds large *1-item* sets from the database. For finding subsequent large item sets it is checked that the item sets found should be greater than the number of columns. We first find candidate item sets from the previous large item sets and then find the subsequent large item sets from the database until all the large item sets are found from the database. In step 4.3 the TBAR algorithm has been modified to store all *L1*s in the knowledge base for subsequent reuse of that information.

V. ARU Algorithm

The ARU algorithm differs from all other update algorithms for association rules as it updates the large item sets in relational databases. So the large 1-item sets are related to a column in a table rather than a transaction in transactional databases. In our case we will find the support for each item set corresponding to a column value in the database.

Inputs

DB=initial database before any updates

db=update portion of the database

DB + db=whole updated portion of the database

L1 DB = large 1-item sets item sets found in DB

attr = attribute in L1 DB

attr.number=attribute number

attr.value=attribute value

attr.count=support of the attribute value

Output

L1 $_{DB+db}$=large 1-item sets in updated database DB+db

ARU ALGORITHM

```
If there is any insertion in the database (Step 5.1)
For L1 DB of attribute attr in database
Get the column number attr.number of the 1-item sets L1 DB
  For all values attr.value from db for the attribute attr.number
```

```
            If the value in the db for the attribute attr.number is
            also in L1DB
                Find support of attr.value in db
                Add support of DB and db
             If the support of DB and db is large in the updated
             database
                Update the support count in the large 1-item sets
             End If
             Else If the value in db is not in L1DB
                Find support of attr.value in db
                    If attr.value is large in db
                        Find support of attr.value in DB
                        Add support of DB and db
                    If the support of DB and db is large in the
                    updated database
                        Update the support count of the
                        attr.value in the large 1-item sets
                    End If
             End if
         UL1 DB + db= updated L1 DB + db for attribute attr
     End for
 Else If no insertions are done in the database
             UL1 DB + db= L1 DB for attribute attr
End If
Generate the item set tree for UL1 DB + db.
Generate all other Lk s from L1 stored in item set tree as in
TBAR Algorithm

Generate association rules from all the Lk s found in DB +
db that are above the minimum confidence threshold

End ARU algorithm
```

The *attr* in the inputs for our algorithm shows us particular attributes that are large in the original database DB. In the step 5.1 we will check to see if there are any insertions in the database, if there are any insertions then all the L1 $_{DB}$ from the knowledge base are reused to find subsequent L_{ks} in DB + db. If there are no updates all L1 $_{DB}$ are taken as the final updated L1s.In subsequent steps these L1s are reused to find all Lk_s from the database.

VI. Experimental Studies

We have checked our algorithm with the TBAR algorithm for 1000 tuples with minimum support threshold from 1 to 5. As shown in Figure 2, ARU algorithm takes much less CPU utilization as compared to TBAR.

In the scale up experiments, we have checked the performance of our algorithm TBAR for 2 % minimum support and with 1000 to 5000 tuples. In Figure 3 it is clear that our algorithm gives linear results in nature, which means that it can be adapted to large databases. Our algorithm is 2.1 to 2.3 times faster than TBAR algorithm.

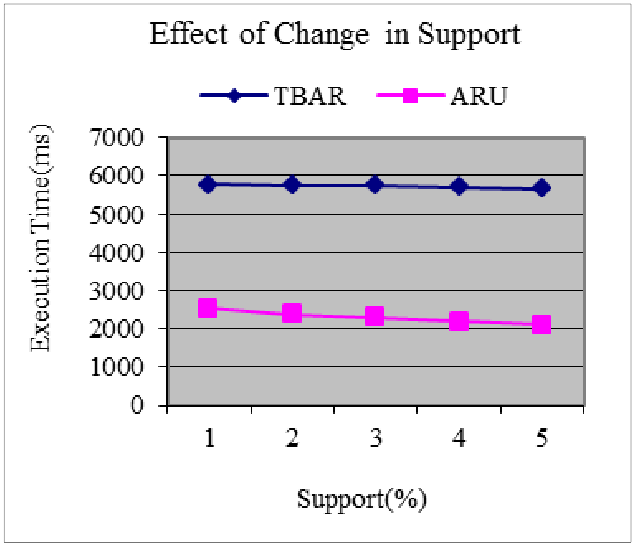


Figure 2. Effect of change in support.

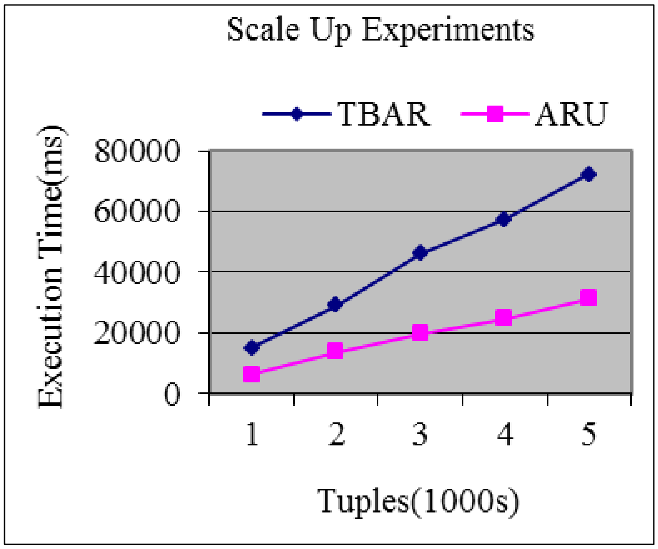


Figure 3. Scale up experiments.

VII. Conclusion

We have presented ARU algorithm, which outperforms the TBAR algorithm. Our proposed algorithm will be able to maintain large items sets by reusing the large item sets found through the initial mining algorithm. Our performance study shows that the proposed algorithm is 2.1 times to 2.3 times faster as compared to the TBAR algorithm. We found the incremental association rules from dynamic databases by employing a new knowledge management technique for relational databases. As a further step our knowledge management technique can be applied to other data mining techniques. Finding association rules from distributed databases is also important area of research.

Acknowledgment

The authors wish to acknowledge the financial support provided by the al Ghurair University.

REFERENCES

[1] R. Agrawal, T. Imielinski and A. Swami, Mining association rules between sets of items in large databases, Proceedings of the 1993 ACM SIGMOD International Conference on Management Of Data, Washington D.C., May 1993.

[2] Dogan and A. Y. Camurcu. “Association Rule Mining form an Intelligent Tutor”, Journal of Educational Technology Systems, Volume 36, Number 4/2007 – 2008, pp 444 – 447, 2008.

[3] D.W. Cheung, J. Han, V.T. Ng and C.Y.Wong, Maintenance of Discovered Association Rules in Large Databases: An Incremental Updating Technique, 1996 International Conference on Data Engineering, New Orleans, Louisiana, February 1996.

[4] D. W. Cheung, V. T. Ng and B. W. Tam, Maintenance of Discovered Knowledge: A Case in Multi-level Association Rules, 2nd International Conference on KDD, Oregon, August 1996.

[5] J.S. Park, M.S. Chen and P.S. Yu, An effective hash-based algorithm for mining association rules, Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, San Jose, California, May 1995.

[6] S. Sarawagi, S. Thomas and R. Agrawal, Integrating Association Rule Mining with Relational Database Systems: Alternatives and Implications, IBM Research Report, 1998.

[7] N.F. Ayan, A.U. Tansel, and E. Arkun. An Efficient Algorithm to Update Large Itemsets with Early Pruning. Proc. of 1999 Int. Conf. on Knowledge Discovery and Data Mining, 1999.

[8] D.Cheung,S.D.Lee and B.Kao.A General Incremental Technique for Updating Discovered Association Rules.Proc.International Conference On Database Systems For Advanced Applications,April 1997.

[9] H. Mannila, H. Toivonen and A.I. Verkamo, Improved Methods for Finding Association Rules, Department of Computer Science, University of Helsinki, Helsinki, Finland, December 1993 (Revised February 1994).

[10] TBAR: An efficient association rule mining for relational databases (1998).

[11] Chang-Hung Lee, Cheng-Ru Lin, and Ming-Syan Chen, Sliding-Window Filtering: An Efficient Algorithm for Incremental Mining, ACM CIKM 2001.

[12] Hipp, U. Güntzer, and G. Nakhaeizadeh. Algorithms for association rule mining — a general survey and comparison. SIGKDD Explorations,2 (1):58—64, July 2000.

AUTHORS PROFILE

Dr. Muhammad Abaidullah Anwar is working as Assistant Professor and Deputy Dean of College of Engineering and Computing in Al Ghurair University, UAE. He received his Doctorate of Engineering with specialization in Object-oriented Databases from Kyushu Institute of Technology, JAPAN in 2001. Since 2001, he has been affiliated with renowned universities in GCC and Pakistan. He has published many papers in International proceeding and journals.

Cloud Computing for Solving E-Learning Problems

N. S. Abu El-Ala
Elec. & Comm. Dept.; Industrial Technical Institute;
Port-Said Technological College; Egypt

W. A. Awad
Math. & Comp. Dept.; Faculty of Science;
Port Said University; Egypt

H. M. El-Bakry
Inf. Systems Dept;
Faculty of Computer and Information Sciences;
Mansoura University; Egypt

Abstract—**The integration of information and communication technologies in education according to the global trend occupied a great interest in the Arab world through E-Learning techniques and put it into the form of services within Services Oriented Architecture Technique (SOA), and mixing its input and outputs within the components of the Education Business Intelligence (EBI) and enhance it to simulate reality by educational virtual worlds.This paper presents a creative environment derived from both virtual and personal learning environments based on cloud computing which contains variety of tools and techniques to enhance the educational process. The proposed environment focuses on designing and monitoring educational environment based on reusing the existing web tools, techniques, and services to provide Browser-based-Application.**

I. Introduction

There is much debate among those interested in education and E-Learning in the Arab world about the advantages and disadvantages of this system of education, and most Arabic academic conference papers presented concerned completely with the same topic which is the advantages and disadvantages and the comparison with the traditional educational systems, so in this paper we will move from E-Learning evaluation to fly in the space of implementation.

The problem is no longer to prove the importance of E-Learning but it is to switch to E-Learning, most educational institutions collide the barrier of foundation when they plan to move to E-Learning systems which require many hardware and software resources.

The educational cloud provides a magic solution to this problem where any educational institution wants to transform its system to E-Learning system. It will have two choices whether to build its own private cloud or to go to a specific service provider to share in a pubic cloud after defining some parameters.

All the users need a host in a data center somewhere in the world, or even multiple data centers scattered around the world, and cloud computing providers deliver common applications online that are accessed from web browsers, also can provide a storage unit to store all learners documents.

No one can ignore that both web 2.0 technologies have changed the nature of the internet from a place to read static pages to an environment that allows end-users to access applications using web browser without purchasing licenses or upgrade hardware, where all software and files hosted in the cloud and accessed by wide range of hardware devices such as mobile phones, computers or PDAs while the internet connection is available.

The following sections focus on how to use cloud computing to enhance the education process specifically in the Arab world. The first section Addresses cloud computing concept and its features, the second section addresses E-Learning environment based on cloud, and the third section compares between existing sites to enhance the existing Arabic E-Learning sites.

II. Cloud Computing.

No one can deny the vital role that cloud plays in developing educational systems recently, but until now there isn't a specific definition for cloud computing.

Simply cloud computing provides anytime / anywhere services that can be accessed from any device in such a way that user is not responsible for where the services or applications are located or how it maintained or updated, all this and more will be the responsibility of cloud computing service providers such as Google and Amazon EC2, etc...

Cloud computing not only saves the money needed for upgrading many labs' hardware or purchase many software licenses but also it relieves the user from periodic maintenance operations. It also provides a high level of security and privacy.

But before transforming your E-Learning system to the cloud, you should identify which services you need and create a service catalogue (list of services you will provide to learners who will access your site) to determine the parameters you will need from service provider, also it will help you whenever you need to update your system. Recent research papers categorize the cloud services into two basic branches as following:

A. IT Services.

It includes all the services related to the infrastructure of the cloud such as physical resources (as storage devices, school servers, and national communication network, etc), and virtual resources that concern with the management and the access of the physical resources.

B. User Services.

It can be divided into two types, generic services (as E-Mail, search, portal, and social networking) and Education services (as e-portfolio, content access and creation).

After determining the required services which serve the curriculum requirements provided by the institution to the students and teachers, a service level agreement will be established between the institution and cloud provider to define the parameters the institution needs.

1) E-Learning Environments.

There are many different educational environments that serve the educational process based on computer and its technologies. For example Web 2.0 technologies which provide teachers with new ways to engage students, and help student to participate on a global level by using the network as a platform for information sharing, interoperability, user-centered design and collaboration on the World Wide Web.

A Web 2.0 site allows users to interact and collaborate with one another in a social media dialogue as creators of user-generated content in a virtual community, which helped the appearance of learning environments such as virtual learning environment and personal learning environment. Before proposing a framework for E-Learning environment with the aid of cloud computing technology, let's know more about the pervious environments.

a) Virtual Learning Environment (VLE):

It is a set of teaching and learning tools designed to enhance a student's learning experience by including computers and the Internet in the learning process where included web-based access to class content, grades, assessments, and other class tools. It is also a social space where students and teacher can interact through threaded discussions or chat.

It also includes students and teacher "meeting" online through a synchronous web-based application. The teacher is able to present lessons through video, PowerPoint, or chatting. The students are able to talk with other students and the teacher, as well as collaborate with one another, answer questions, or pose questions. They can use the available tools through the application to virtually raise their hands, send messages, or answer questions on the screen given by the teacher.

b) Personal Learning Environment (PLE):

The expression does not refer to a specific service or application but rather to an idea of how learners achieve their learning goals. PLE provides learners with support in managing their content and communication with peers in the process of learning by dividing them into groups for discussions, providing context, and il¬lustrating processes.

PLE provides a suitable environment to practice social skills. There are many types of PLEs that are classified based on their architecture such as PLEX or web-based with loosely joined web services such as ELGG or are classified based on their platform like facebook. Another approach of PLEs is based on their pedagogical approach that serves formal and informal learning process.

c) Mash-up Personal Learning Environments (MUPPLE):

This environment is a mixing between the previous environments where they allow learners to build their own personal learning environment by composing web-based tools, get involved in collaborative activities, share their designs with peers, and adapt their designs to reflect their experience in the learning process.

After identifying the E-Learning environments and recognizing the features of each environment, we will move to the following section which is concerned with the outline a framework for E-Learning environment with aid of cloud technology as shown in Fig.1.

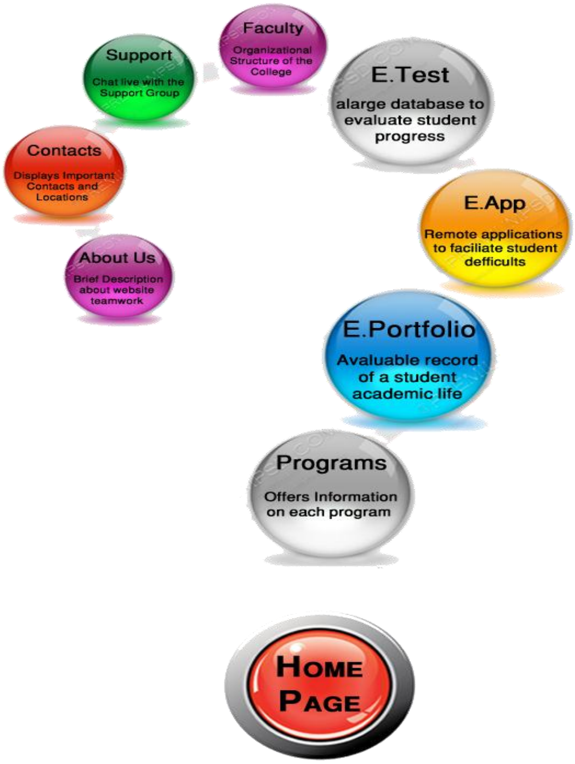

Faculty
Organizational Structure of the College
Support
Chat live with the Support Group
E.Test
alarge database to evaluate student progress
Contacts
Displays Important Contacts and Locations
E.App
Remote applications to faciliate student defficults
About Us
Brief Description about website teamwork
E.Portfolio
Avaluable record of a student academic life
Programs
Offers Information on each program
Home Page

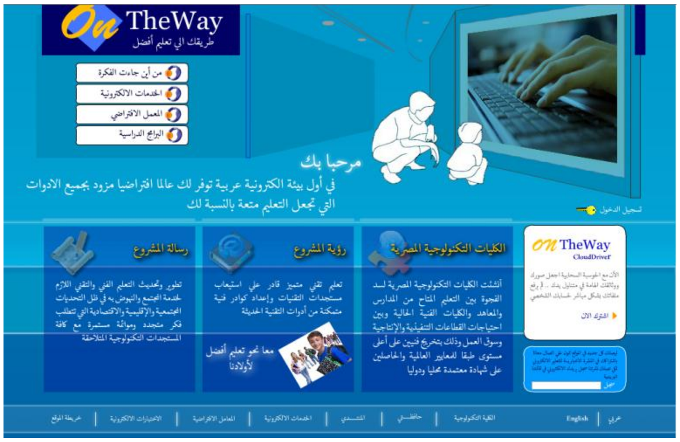
Fig.1 E-Learning Framework

2) Cloud-based E-Learning System.

This section concerns with designing web-based E-Learning system that contains various social tools, smart agents and interactive environment of web 2 techniques uploaded to cloud as shown in Fig. 2. The system has three major parts. The first part addresses the web-based Course Management System (CMS) which is managed by the web server to register learner to access course materials that are provided and maintained by teachers.

The second part is a PLE which provide various tools and services to help learners in building their own environment. The third part addresses building an online virtual computing lab as shown in Fig. 3, Fig. 1 Site Map. providing a remote access service that allows the leaner to reserve a computer with a desired set of applications such as Photoshop, Packet Tracer, AutoCAD, and many others Linux and numerous windows environments and remotely access it over high-speed internet connection.

All of this and more shown through an attractive easy-to-use interface represented in a OnTheWay web interface which achieves a set of criteria to facilitate the students' tasks.



مرحبا بك

في أول بيئة الكترونية عربية توفر لك عالما افتراضيا مزود بجميع الادوات التي تجعل التعليم متعة بالنسبة لك

رسالة المشروع

تطوير وتحديث التعليم الفني والتقني اللازم لخدمة المجتمع والنهوض به في ظل التحديات المجتمعية والإقليمية والاقتصادية التي تتطلب فكر متجدد وموائمة مستمرة مع كافة المستجدات التكنولوجية المتلاحقة

رؤية المشروع

تعليم تقني متميز قادر علي استيعاب مستجدات التقنيات وإعداد كوادر فنية متمكنة من أدوات التقنية الحديثة

معا نحو تعليم أفضل لأولادنا

الكليات التكنولوجية المصرية

أنشئت الكليات التكنولوجية المصرية لسد الفجوة بين التعليم المتاح من المدارس والمعاهد والكليات الفنية الحالية وبين احتياجات القطاعات التنفيذية والإنتاجية وسوق العمل وذلك بتخريج فنيين على أعلى مستوى طبقا للمعايير العالمية والحاصلين على شهادة معتمدة محليا ودوليا

الآن مع الحوسبة السحابية اجعل صورك ووثائقك الهامة في متناول يدك .. قم برفع ملفاتك بشكل مباشر لحسابك الشخصي

Fig. 2 OnTheWay Web Site, Homepage.

- OnTheWay achieves the learning objectives accredited curriculum department of the Ministry of Higher Education and teacher-defined course goals.

- OnTheWay provides the means to deliver high quality digital learning applications fully integrated into any course. Learners learn by access text and multimedia content.

- OnTheWay provides both teachers and learners the ability and storage area to upload all their documents, projects, homework and photos at their fingertips.

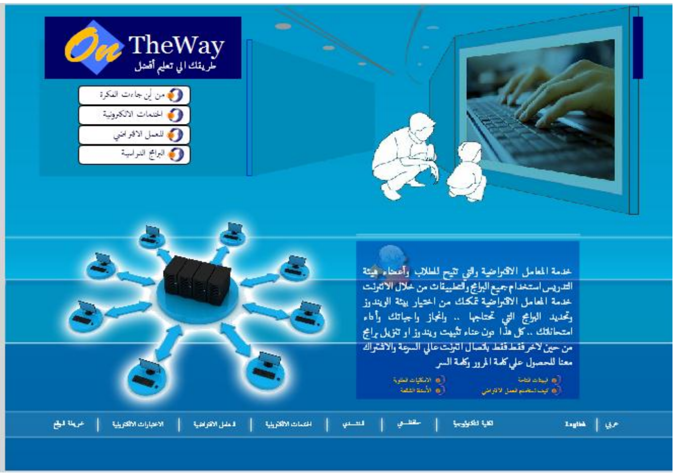
- OnTheWay provides the technology, tools and professional development that facilitate the students' task and help them to share and present their ideas, thinking and learning by using OnTheWay forums.

- OnTheWay provides the technology, tools that helps students to access their virtual computing lab to implement their educational activities by providing a number of Windows and Linux environments and set of suggested programs.

- OnTheWay provides the students with online test which provide various questions on various topics to improve the students' skills.

- OnTheWay provides learners the ability to follow up their performance level by assignment report, progress report, working portfolios, and projects Evaluation.

The user interface is friendly, and accessible.


OnTheWay
طريقك الي تعلم أفضل
من أين جاءت الفكرة
الخدمات الالكترونية
للعمل الافتراضي
البرامج الدراسية
الإمكانيات
الأسئلة الشائعة
تعليمات عامة
كيف تستخدم المعمل الافتراضي
عربي
الكلية التكنولوجية | حافظتي | المنتدي | الخدمات الالكترونية | المعمل الافتراضية | الاختبارات الالكترونية | خريطة الموقع

خدمة المعامل الافتراضية والتي تتيح للطلاب وأعضاء هيئة التدريس استخدام جميع البرامج والتطبيقات من خلال الانترنت خدمة المعامل الافتراضية تمكنك من اختيار بيئة الويندوز وتحديد البرامج التي تحتاجها .. وانجاز واجباتك وأداء امتحاناتك .. كل هذا دون عناء تثبيت ويندوز او تنزيل برامج من حين لآخر فقط فقط باتصال انترنت عالي السرعة والاشتراك معنا للحصول علي كلمة المرور وكلمة السر

Fig. 3 OnTheWay Web Site, Virtual Computing Lab (VCL) Page.

IV. Conclusion

In this paper we tried to prove that cloud computing changed E-Learning future systems. A wide world of knowledge and tools now is available to Arabic teachers and learners through cloud based services all the time and accessed from anywhere, from any device.

Reference:

[1] .Paul POCATILU; "Cloud Computing Benefits for E-Learning Solutions"; Economics of Knowledge; Vol. 2; Issue 1; 1Q 2010; pp. 9-14.

[2] DeCoufle B.; "The Impact of Cloud Computing in Schools"; The Datacenter Journal; http://datacenterjournal.com/content/view/3032/40/; July 2009.

[3] Pocatilu P., Boja C.; " Quality Characteristics and Metrics related to M-Learning Process"; Amfiteatru Economic; Year XI; No. 26; June 2009.

[4] Pocatilu P., Boja C.; "Cloud Computing Incidents Database"; wiki.cloudcommunity.org/wiki; 2010.

[5] Cena F. Farzan R. Lops P.; "Web 3.0: Merging Semantic Web with Social Web"; Proceedings of the 20th ACM conference on Hyper-text and hypermedia; HT'09; June 29– 1 July; 2009.

[6] Mohammed Al-Zoube; "E-Learning on the Cloud"; International Arab Journal of E-Technology; Vol. 1; No. 2; June 2009.

[7] WHITE PAPER, Intel® World Ahead Cloud Computing; "The Education Cloud: Delivering Education as a Service"; 2010.

[8] http://www.educause.edu/eli; EDUCAUSE Learning Initiative; advancing learning through it innovation ; "7 things you should know about.. Personal Learning Environments"; May 2009.

[9] Fridolin Wild, Felix Mödritscher and Steinn Sigurdarson; "Designing for Change: Mash-Up Personal Learning Environments"; www.elearningpapers.eu; July 2008.

[10] http://www.fullerton.edu/VCL/ index.asp; July 2012.

[11] http://www.vcl.ncsu.edu/; May2012.

Optimizing the Performance Evaluation of Robotic Arms with the Aid of Particle Swarm Optimization

K Shivaprakash Reddy
Department of Mechanical Engineering,
The Oxford College of Engineering,
Bommanahalli, Hosur Road,
Bangalore – 560068, Karnataka, India

Dr.PVK Perumal
Department of Aeronautical Engineering,
M.A.M School of Engineering,
Trichy – Chennai Trunk Road,
Siruganur, Trichy – 621105, Tamil Nadu, India

Dr.B. Durgaprasad
Department of Mechanical Engineering,
JNTU College of Engineering, Jawaharlal Nehru
Technological University,
Anantapur – 515002, Andhrapradesh, India

Dr.M.A Murtaza
Department of Mechanical Engineering,
The Oxford College of Engineering,
Bommanahalli, Hosur Road, Bangalore – 560068,
Karnataka, India

Abstract— **In this modern world, robotic evaluation plays a most important role. In secure distance, this leads the humans to execute insecure task. To acquire an effective result, the system which makes the human task easier should be taken care of and the holdup behind the system should be eradicated. Only static parameters are considered and such parameters are not enough to obtain optimized value in existing work. For consecutively attaining optimized value in our previous work, we focused on both static and dynamic parameters in the robotic arm gearbox model. Now, a genetic algorithm is utilized and the result obtained is greater than the existing work. On the other hand, to attain an effective result the genetic algorithm itself is not enough since it takes massive time for computation process and the result obtained in this computation is not as much closer to the true value. By eliminating all those aforementioned issues, a proper algorithm needs to be utilized in order to achieve an efficient result than the existing and our previous works. In this paper, we anticipated to suggest a Particle Swarm Optimization technique that reduce the computation time as well as make the output result as much closer to the true value (i.e.,) experimentally obtained value.**

Keywords-Particle Swarm Optimization; Robotic arm gear box; Static& Dynamic parameters.

I. INTRODUCTION

In the physical world, robots are physical agents that attain tasks by manipulation. In general, to sense the environment and effectors to claim physical forces on it robots are equipped with sensors. In the automation process, industrial robots play a most important role in grinding method. The majority of the grinding robots work in a inhibited environment, where instantaneous position and force control is vital [1]. Manipulators, Mobile robots, and Humanoid robots are the three main categories [2]. To enhance product quality and safety, while minimizing costs and processing time, robots find concentrated applications in factories. The robot model depends on the inertia, mass, and center of mass of each link [3]. Tele-manipulators and the capacity of numerical control of machines are the two prominent technologies in which robotics are based on. Tele-manipulators are remotely controlled machines that often hold an arm and a gripper. According to the instructions given by the humans through his/her control device, the movements of arm and gripper will take place. With respect to a given coordinate system, numeric control allows controlling of machines very accurately [2].

A. Types of Robot

1) Mobile Robots

A special group of effectors for locomotion, such as wheels, tracks, and legs are used by the mobile robots. The differential drive contains two independently actuated wheels, one on each side. When the movement of both wheels is at equal velocity the robot travels in a straight line. The robot turns on the spot if they move in opposite directions. The development of mobile robots was motivated by the desire to automate transportation in production processes and autonomous transport systems. New types of mobile robots have been created recently like insectoid robots with several legs modeled after examples nature gave us or independent robots for underwater usage.

2) Hard working Robots

Mostly, in areas of difficult toil robots have been used to replace human workers, which are structured enough for mechanization, like assembly line work in the automobile industry (the classical example) or harvesting machines in the agricultural zone. A few existing examples apart from the assembly robot are Melon harvester robot, Ore transport robot for mines, robot that removes paint from big ships and a robot that creates high precision sewer maps. If robot is used in a proper environment, then it can work faster, cheaper and more exact than human beings.

3) Transporters

Most autonomous transport are widely in use since the robots still desires environmental changes to find their way. But, designing a robot that can navigate using natural landmarks is probably an end to science fiction. Examples of currently available transporters are (1) Container transporters employed to load and unload cargo ships, (2) Medicine and food transport systems in hospices, and (3) Autonomous helicopters employed to transport goods to distant areas.

4) *Insensible Steel Giants*

Since robots can be easily protected against hazardous environments and are adequately replaceable, they are used in perilous, toxic or nuclear environments. In some places, for cleaning up a mess robots have been used. For example, in Chernobyl disaster, to clean up the nuclear waste, and also robots are employed to clean grenades and mines all around the world robots have helped. Moreover, robots are sent to Mars and into the depth of the oceans. They can also investigate deep-set ships and can walk on the craters of active volcanoes.

5) *Servants and Toys*

In our world, robots may not yet be a common sight, but in several places we already meet and used them. A lot of modern toys like the Sony Aibo are spoiling the today's children's life. To help the older people robots are generally developed to have a better and more secure life. Today, in the name of toys or household helpers they begin to come with us [2].

6) *Industrial Robots*

Industrial robots have been entrenched in the manufacturing area, used for performing the tasks such as stacking, casting, painting, sorting, welding, component soldering and more for more than thirty years. This use framework highlights the core value proposal of an industrial robot: performing tasks incessantly and precisely in work environments and scales difficult for humans. For performing operations swiftly, continually and precisely, industrial robots are developed. In the manufacturing industry, operating in relatively static environments and in large numbers, they have a long legacy. To enhance the security and efficiency as well as to decrease the environmental impact, the oil and gas industry suggest the use of industrial robotics. New developments in regions that are difficult or unsafe for humans to work in could be easily handled and maintained by remotely-controlled industrial robots [4].

B. *Gear-box*

To change the speed from the low rotating rotor to the high rotating generator a gearbox is used [5]. In their transmissions robotic systems that need to present high torques at the end effectors typically contain high reduction gears, causing some gear-specific friction components to appear, such as position dependent friction. It produces a periodic waveform friction with the frequency by which the two teeth match, since this force happens once a pair of teeth comes together. Thus, as position dependent friction, it has been always considered. The parameters utilize for considering the functioning of gearbox are (1) Direction of turning, (2) Relative speed, (3) The number of revolution, (4) Mechanical advantages concerning the principles of power and speed [6].

Friction takes place along the off line-of-action direction, which lies orthogonal to the line of action, when the power is transmitted along the line of action direction. The major reason for this friction is that the teeth slide together as an alternative of rolling absolutely. Hence, in gearboxes meshing friction is a source of uselessness. The power and load in gear trains are transmitted along the line of action. For small shaft displacements along the line-of-action direction, the relative reduced rigidity of shaft support ball bearings may be responsible, which would guide to torque oscillations and this is called position dependent friction. Though, the meshing friction force in gear teeth is transmitted in the off line-of-action direction. The friction coefficient between the gear teeth significantly depends on lubricant properties, and it decreases as the relative sliding velocity between gear teeth increases.

C. *Types of Gears*

The amplitude of the oscillation caused by meshing friction also depends greatly on the gear type. Between teeth the working principle of spur or helical gears is rolling. Therefore, in some cases the meshing friction could be small. However, the working principle of other gear types like worm gears, is approximately pure sliding friction. The lubricant film is insufficient to prevent contact between asperities, and friction becomes inflated, when the speed is low. The lubricant film will become adequate to reduce friction, only if speed increases [7].

D. *Common Gearbox Parameters*

Some of the common gear-box parameters are

1) *Viscous friction coefficient*

Viscous friction between two surfaces that have relative motion between them relies on dimensional parameters such as contact area and approval between the two surfaces, and also relies on fluid properties, such as fluid specific gravity and viscosity. Between the two meeting surfaces, viscous friction is inversely proportional to the clearance [8].

2) *Coulomb friction coefficient*

In contact with each other, coulomb friction is a basic measurement of the friction force that exists between two dry surfaces. The coulomb friction coefficient is a static force, which is to some extent higher than motive force when two materials are at rest whereas in contact with each other. For several simple, pure materials and is given as a unit-less number, this coefficient of friction is distinguished. The coefficient of friction for wood against concrete is 0.62, for polystyrene against steel is 0.3 to 0.35, and for steel against Teflon is 0.04 for dry surfaces. To compute the force required to conquer static friction, called as the friction force these numbers are utilized, by multiplying the coefficient of friction times the normal force. The normal force is the mass of the materials times' gravitational pull, with vector calculations added in if the two surfaces are moving up or down an incline against gravitation pull, or towards it [9].

3) *Striebeck friction coefficient*

As a function of a dimensionless lubrication parameter ηN/P, Striebeck systematically studies the variation of friction between two liquid lubricated surfaces, where η denotes the dynamic viscosity, N represents the speed i.e., revolutions per minute of a bearing, and P represents the load anticipated on to the geometrical surface [10].

4) *Friction smoothness coefficient*

Friction is the resistance that an object encounters when moving over another (OED). Since the sandpaper exerts more frictional resistance, it is fast and effortless to drag an object over glass than sandpaper. It was implicit that a surface does not use any frictional force if it is "smooth", in many situations. However, this wouldn't be the case in real life. A "rough" surface is one that provides some frictional resistance [11].

5) Total moment of inertia

In traditional mechanics, moment of inertia, also called mass moment of inertia, rotational inertia, polar moment of inertia of mass, or the angular mass, (SI units kg•m²) is a measure of an object's resistance to any change in its state of rotation. It is the inertia of a rotating body corresponding to its rotation. The moment of inertia plays much the same role in rotational dynamics as mass does in linear dynamics, depicting the relationship between angular momentum and angular velocity, torque and angular acceleration, and numerous other quantities. The symbol ‘I’ and sometimes ‘J’ are often used to represent the moment of inertia or polar moment of inertia [12].

In this paper, by considering the parameter values our primary intention is to decrease the variance occurring between the theoretical value and the practically obtained experimented value. We obtain the optimized parameter value, by utilizing genetic algorithm, which in case positively reduce the error occurred and also the result obtained after applying the optimized parameter value almost bring the intended value and experimentally obtained value.

II. Related Works

The most noticeable features of robotic applications are heterogeneity. With a variety of hardware and software that must be integrated efficiently to develop applications that not only satisfy classic robotic requirements but also software engineering aspects, large robotic projects engage numerous different researchers. However, either they do not cope with such heterogeneity or do not embrace specific robotic requirements in most prior solutions to this problem. In 2008, Juan-Antonio Ferna´ ndez-Madrigal et al. [13] have proposed a framework for the implementation of heterogeneous robotic software via a software engineering technique. The BABEL development system, the main phases of the application lifecycle such as, design, implementation, testing, and maintenance are intended to cover when unavoidable heterogeneity conditions are present. For designing and implementing different real robotic applications that employs various programming languages (C, C++, JAVA), execution platforms (RT-operating systems, MS-Windows, no operating system at all), communication middleware (CORBA, TCP/IP, USB), and also various hardware components (PC, microcontrollers, and a wide variety of sensor and actuator devices in mobile robots and manipulator arms), the potency of proposed system have been revealed by its support.

In 2008, Sungho Jo [14] has proposed a biologically inspired robotic model. This model is developed combining modified feedback error learning, an unsupervised learning, and the viscoelastic actuator system. They are integrated in order to drive adaptive arm motions, and also discussed the potential efficacy of a biomimetic design of robot skill. With the cerebellar adaptation, the unsupervised learning, the synergy network adaptation, and the viscoelastic system of the muscles the feedback error learning was reliable. To control the redundant actuators efficiently, the proposed model has used a feed forward adaptive approach in the low dimensional control space and an adaptive synergy distribution. With six muscular actuators in the gravitational field, the amalgamation of the two adaptive control approaches has been tested by controlling a two-link planar robot arm. To make smooth, human-like motions, the simulation-based study has shown that the control method can adapt the robot arm motions swiftly and robustly.

Over their rigid counterparts, flexible robot manipulators have abundant advantages. They have increased payload-to weight ratio, they operate at higher speeds, employ less energy and smaller actuators, and they are secure during interaction with their environments. Conversely, light design along with external effects result in components which can oscillate with extreme amplitudes. These oscillations cause deviation from the desired path and long idle periods between tasks in order to perform the intended operation securely and precisely. In 2008, Abdullah Ozer et al. [15] have examined the efficiency of a vibration control method for a two-link flexible robotic arm. Variable stiffness control (VSC) method has been employed to control the excessive oscillations. Due to its dissipative nature, the method was stable and it was relatively insensitive to significant parameter changes and suitable to be implemented on existing robots. Their research considers that the source of the flexibility was either the joints or the links or both. Simulation results have been presented to exhibit the flexibility of the proposed control method. Experiments have been conducted on a laboratory prototype and the results have been presented to prove the validity of simulations.

A human performs a variety of adroit movements by adjusting the dynamic characteristics of his/her musculoskeletal system according to a task involved. By mechanical impedance parameters such characteristics of human movements can be represented. There is a chance that human skillful strategies can be included into robot motion control, if the regulation mechanism of human impedance properties during the task can be clarified and modeled. In 2008, Toshio Tsuji et al. [16] have studied the human hand impedance in preparation for task operations, the so-called “task-readiness impedance”, in a virtual ball-catching task. For contact tasks by computer simulations using measured task-readiness impedance they have also discussed a bio-mimetic impedance control of robotic manipulators.

In 2009, Freeman et al. [17] have developed an experimental test facility. This facility was developed for the use by stroke patients in order to enhance the sensory-motor function of their upper limb. Subjects have been seated at the workstation and their task is to continually follow reaching trajectories that are projected onto a target above their arm. To perform this, they used a voluntary control with the addition of electrical stimulation mediated by advanced control methods applied to muscles in their impaired shoulder and arm. The particulars regarding the design of workstation and its periphery systems have been given, together with a depiction of its use during the healing of stroke patients.

In 2009, Yunquan Sun et al. [18] have performed robotic belt grinding operations. This was done by increasing a work piece to the end effecter and imposing it to move along a route while maintaining contact with the belt grinding wheel. To give a smooth finish on the work piece, a constant contact force throughout the grinding process was essential, but it was tricky to maintain this force due to a multitude of installation, manipulation, and calibration errors. The proposed methodology for robotic belt grinding has been described, which mainly concentrates on system calibration and force control to enhance grinding performance. For each step of the proposed method have been shown, the overall theory has been explained and experimental results of turbine blade grinding.

Antagonistic Driven Compliant Joints (ADCJs) are object, drawn substantial attention in current robotics research, representing one of the most extensively applied solutions. This was developed to develop human-like and safe joints for human-robot communication. In 2010, Nicola Vitiello et al. [19] have proposed a sensor less torque control technique, appropriate for ADCJs actuated robots. Off-line characterization of the flexibility of the actuation units, defined by the force–elongation curve and online estimation of the force exerted by each actuation unit, through a direct measure of the joint angle, and of the ‘‘resting position” of each actuation unit are the two steps followed in the proposed technique. To develop two autonomous force controllers, the proposed force estimation technique has been employed, with no need of additional torque sensors that can be then fused to control the resulting joint torque. Over the shoulder and the elbow ADCJs of the 2-link 2-DOFs planar robotic arm NEURARM, the performance of the proposed torque control has been analyzed. The technique was proved to work effectively, achieving better performances on the test platform, and represents a suitable alternative to modern sensor-based torque controls.

In 2011, Celso De La Cruz et al. [20] have developed a dynamic model of a robotic wheel chair. This was developed considering a lateral deviation of the center of mass. To design a tracking and positioning adaptive control for the robotic wheel chair, the Lyapunov and input/output stability theories have been utilized. Regarding to its matrices and parameters, properties of the dynamic model have been exhibited. A filter has been engaged to obtain a closed loop equation that permits designing of adaptive control law. Consequently, to improve the adaptive control in the sense of eliminating parameter drift a projection algorithm has been used. Experimental results have shown the better performance of the adaptive control.

In 2011, C.M. Wronka et al. [21] have proposed a dynamic model of a robotic manipulator mounted on a moving base. This was designed by means of the Euler–Lagrange technique. In the dynamic equations, it is assumed that the base inertia was large enough not to be affected by the manipulator motion and hence can be treated as a time-varying parameter. To a Mitsubishi PA10-6CE robotic manipulator mounted on a 2-DOF platform, the presented derivation has been applied. By comparing simple closed-loop control results of the simulated model with experimental data from the manipulator mounted on the platform the model has been evaluated.

In real environments, achieving manipulation tasks interactively necessitates a high level of precision and stability. One must provide the robot with the skill to react quickly to abrupt changes in the environment at the same time when one cannot assume a fully deterministic and static environment. In 2012, Ashwini Shukla et al. [22] have recorded the kinematics of arm and fingers of human subjects. This was done during unperturbed and perturbed reach and grasp motions. After the onset of the motion the target’s location has been changed abruptly in the perturbed demonstrations. Between the hand transport and finger motions, data has shown a strong combination. To seamlessly and rapidly adapt the finger motion in coordination with the hand posture, they theorize that the coupling enables the subject. A coupled dynamical system based controller has been proposed, whereby two dynamical systems driving the hand and finger motions have been coupled to provide their robot with the capability. For reach-to-grasp motions the proposed method has provided a compact encoding that ensures fast adaptation with zero latency for re-planning. They have proved that this coupling ensures smooth and ‘‘human-like’’ motions from the simulation performed on the real iCub robot.

In 2012, Hassan Azarkish et al.[23] have presented the performance of the particle swarm optimization and the genetic algorithm compared as a typical geometry design problem. From a given fin volume, the plan maximizes the heat transfer rate. The analysis presumes that a linear temperature sharing the length of the fin. Using the B-spline curves the fin profile generated and restricted by the alteration of control point coordinates. An inverse method applied to find the appropriate fin geometry yield the linear temperature distribution along the fin corresponds to optimum design. The numbers of the populations, the count of iterations and time to convergence measure efficiency. For geometry optimization, results show that the particle swarm optimization is most competent.

III. Particle Swarm Optimization Based Optimized Parameter Value

A. Generation of Chromosomes

Initially, numbers of chromosomes are generated randomly. Each chromosome has number of genes. The randomly generated chromosomes can be determined as,

$$C_i^{(k)} = \left\{c_0^{(k)}, c_1^{(k)}, c_2^{(k)}, \cdots, c_{N-1}^{(k)}\right\} \qquad (1)$$

Here, C_i - No of Chromosomes, c_j - No of genes in Chromosomes, and every ithvalue has its own range, which should be different from the range of other genes, and the minimized range will give better result when compared to the existing work.

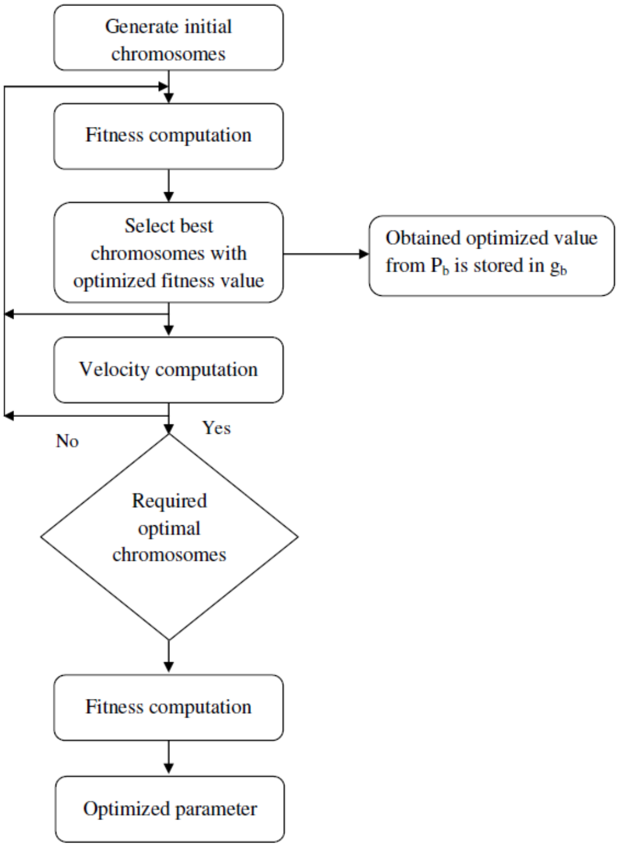
Fitness computation

$$E_{err} = (Ev - Mv)$$

$$S_{el} = \sum_{x=1}^{N_x} E^x err$$

$$m = \frac{S_{el}}{N_x}$$

If min(m)
End If


Generate initial chromosomes
Fitness computation
Select best chromosomes with optimized fitness value
Obtained optimized value from P_b is stored in g_b
Velocity computation
No
Yes
Required optimal chromosomes
Fitness computation
Optimized parameter

To compute the fitness, all the obtained chromosomes and the corresponding parameter values are taken from the eqn (1). Here, error minimization is the fitness function, which is shown below.

Where, Ev – Experimental value; Mv – model value ; - Error Element ; - sum of error Elements ; - No of Elements; - Mean of error elements, which is also said to be Fitness. From the above pseudo code, select number of chromosomes to be applied with the genetic operations for velocity computation.

Objective Function

$$\dot{x}(t) = \begin{bmatrix} x_3 - x_4 \\ x_4 - x_5 \\ \frac{1}{I_m}(-T_{st}(x_1) - D_g(x_3 - x_4) - T_{fr}(x_3) + u) \\ \frac{1}{I_g}(T_{st}(x_1) + D_g(x_3 - x_4) - k_s x_2 - D_a(x_4 - x_5)) \\ \frac{1}{I_a}(k_s x_2 + D_a(x_4 - x_5)) \end{bmatrix} \qquad (2)$$

$$I = I_{ai} + I_{mi} + I_{gi} \qquad (3)$$

I_{mi} = Moment of inertia of motor

I_{ai} = Moment of inertia of arm

I_{gi} = Moment of inertia of gear box

T_{stq} = Spring Torque

D_{pa} = Damping Parameters of arm

D_{pg} = Damping parameters of gearbox

k_{ss} = Stiffness of the second spring

$$T_{fr}(x_3) = C_v x_3 + (C_c + C_{cs}\operatorname{sec}h(\sigma x_3))\tanh(\tau x_3) \qquad (4)$$

=Viscous friction coefficient

= Coulomb friction coefficient

= Striebeck friction coefficient.

= Friction torque

$$T_{stt}(x_1) = k_{gb1}x_1 + k_{gb3}x_1^3 \qquad (5)$$

= Gearbox stiffness parameter 1

= Gearbox stiffness parameter 2

= Spring Torque

$$E_r(\theta) = \frac{1}{N}\sum_{t=1}^{N} e^2(t,\theta) \qquad (6)$$

Is said to be the Error value

$$e(t,\theta) = y(t) - y(\hat{t},\theta) \qquad (7)$$

If , then the simulated output of the model is obtained with the input u (t) and without e (t) for the current parameter vector. The criterion (6) is minimized by an iterative numerical search algorithm, which involves simulation of the system for different values of . [24]

$$\min_{\theta} E_r(\theta) = \frac{1}{N}\sum_{t=1}^{N} e^2(t,\theta) \qquad (8)$$

B. *Velocity computation*

Particle Swarm Optimization has two significant operators: velocity update and position update. Particle Swarm Optimization utilizes numerous particles to investigate minimum values of an objective function.

Every particles move in a certain search space with a velocity. Based on the current velocity of each iteration and, a newly obtained particle are calculated. The velocity and position of the particle will be updated according to the equations given below.

$$v_m = v_c + l_{f1}r_1(p_b - p) + l_{f2}r_2(g_b - p) \qquad (9)$$

$$p_m = p + v_m$$

v_c = Current velocity vector

p = Current position vector

p_m = Modified position vector

r_1 =Random parameter1 (0 - 1)

r_2 = Random parameter2 (0 - 1)

l_{f1} = Learning Factor1

l_{f2} = Learning Factor2

p_b =Previous best (p_{best})

g_b =Best value tracked by PSO (g_{best})

C. Selection of best chromosomes

After the completion of number of iterations, the best chromosome is chosen from the obtained chromosomes. Here, the best chromosome is one having least error value. Subsequently, the genes of best chromosome are arranged in the rising order and the chromosome that has least error is preferred as the best gene.

From the above process, we have obtained the optimal fitness value i.e., eqn (8). By utilizing this fitness value, we can obtain the value similar to that of the experimental value [24].

IV. Result And Discussion

To obtain the optimized fitness value we are utilizing particle swarm optimization which is more preferable for optimization process here we use optimize fitness value for four sets of experiment (i.e.) Estimation, Validation1, Validation2, Validation 3.which gives more accurate value when compare it with the standard value as well as the value which obtain from our previous technique genetic algorithm.

In table I shows the optimized fitness value in the proposed column which clearly shows that the proposed optimized fitness value is greater than that of existing fitness value and our previous work, the corresponding graph show below figure. 1 clearly explains that our proposed work is more closely near to the experimental value. The standard and existing data are obtained from [25].In the figure 2,3,4,5 explains the general model graph for standard and the existing work amid blue color marked is how much occur in the black color shaded graph. The remaining portion shows that Existing work is lag with the standard value.

TABLE I. Comparison of Existing fitness value, First work Fitness Value and proposed fitness value

Experiments	Existing Fitness Value	1st work Fitness Value	Proposed Fitness Value
Estimation	72.63	78.7	85.14
Validation 1	73.22	79.2	85.56
Validation 2	8.887	37.5	63.6
Validation 3	95.86	98	98.36

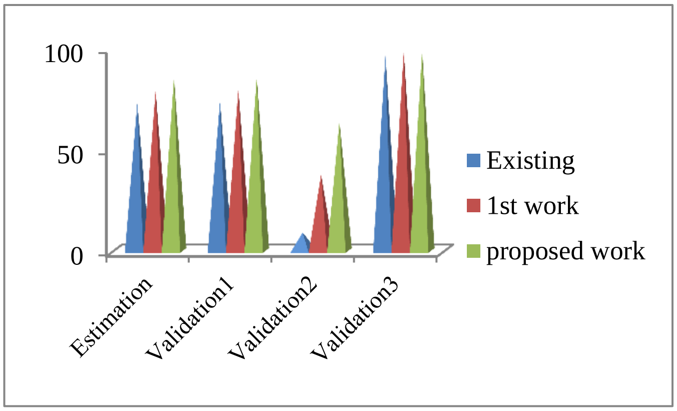


Figure 1. (Fitness comparison)

Graph obtained for Standard and Existing Experimental value for four sets of Experiment

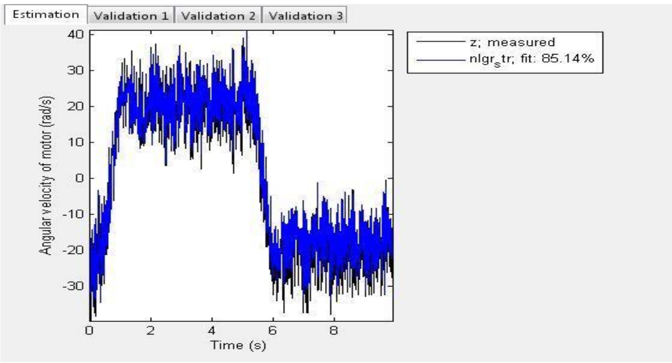


Figure 2. (Estimation)

In Existing work they attain 72.63% of fitness value when compare to that of standard value. (i.e.) the blue color occur only 72.63% in the black shaded part in the graph. In our previous work we attain 78.7%. In this proposed work we attain 85.14% which is greater than that of Existing work and also our previous work.

In Existing work they attain 73.22% of fitness value when compare to that of standard value. (i.e.) the blue color occur only 73.22% in the black shaded part in the graph. In our previous work we attain 79.2%. In this proposed work we attain 85.56% which is greater than that of Existing work and also our previous work.

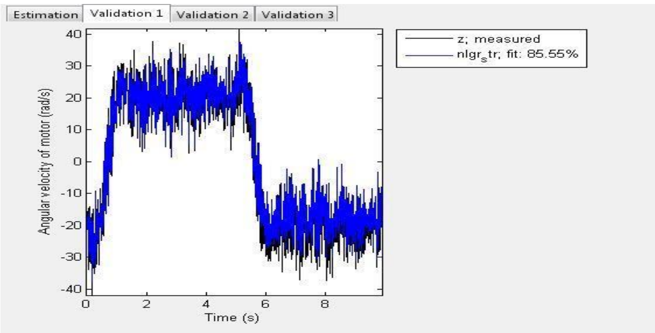


Figure 3. (Validation 1)

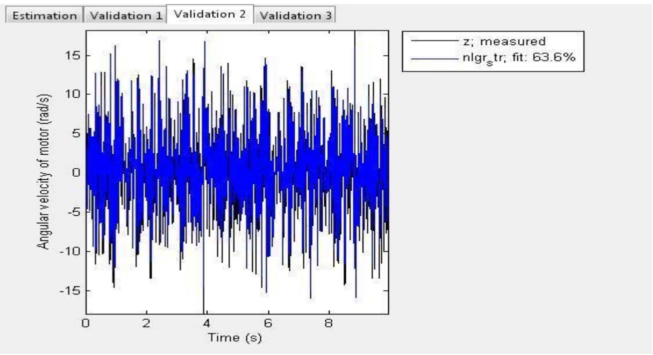


Figure 4. (Validation 2)

In Existing work they attain 8.887% of fitness value when compare to that of standard value. (i.e.) the blue color occur only 8.887% in the black shaded part in the graph. In our previous work we attain 37.5%. In this proposed work we attain 63.6% which is greater than that of Existing work and also our previous work.

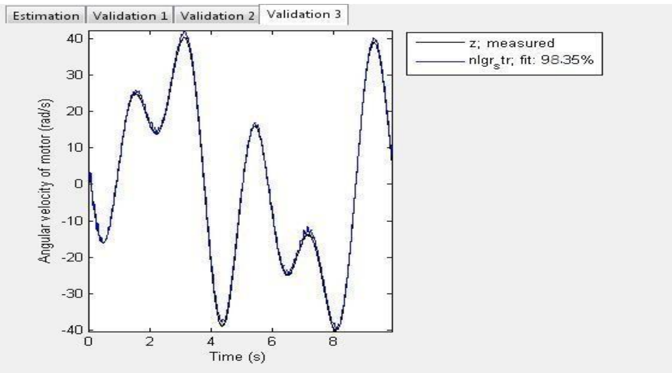


Figure 5. (Validation 3)

In Existing work they attain 95.86% of fitness value when compare to that of standard value. (i.e.) the blue color occur only 95.86% in the black shaded part in the graph. In our previous work we attain 98%. In this proposed work we attain 98.36% which is greater than that of Existing work and also our previous work.

CONCLUSION

The proposed technique was robust and is used to predict the optimized value from the generate chromosomes. In general there is some conflict in experimentally obtained value and numerically obtained value. In order to eradicate those conflicts we have to optimize the parameter value and minimize the error occur in those equation, which certainly gives the optimized value which is closely nearer to that of experimentally obtained value. The obtained optimized value should certainly improve the flexibility of robotic arm.

REFERENCE

[1] Saeid Nahavandi, Mohammad Jashim Uddin, Yasuo Nasu, Hieu Trinh, Mozafar Saadat, "Automated robotic grinding by low-powered manipulator" ,Robotics and Computer-Integrated Manufacturing, Vol.23, pp.589-598,2007

[2] Tim Niemueller and Sumedha Widyadharma,” Artificial Intelligence – An Introduction to Robotics”, AI-Robotics, July 8, 2003

[3] Basilio Bona and Aldo Curatella,” Identification of Industrial Robot Parameters for Advanced Model-Based Controllers Design”, Proceedings of the 2005 IEEE International Conference on, Robotics and Automation (ICRA), pp 1681 – 1686, 2005.

[4] Clint Heyer, ”Human-Robot Interaction and Future Industrial Robotics Applications”, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2010

[5] Svend Gade, Richard Schlombs, Christoph Hundeck and Christian Fenselau,” Operational Modal Analysis on a Wind Turbine Gearbox”, Conference & Exposition on Structural Dynamics, 2009

[6] Joan M. Chambers, Mike Carbonaro and Hana Murray,” Developing conceptual understanding of mechanical advantage through the use of Lego robotic technology”, Australasian Journal of Educational Technology, v24 n4 p387-401 2008

[7] S.E. Hodges and R.J.Richards, ”Look and Learn: Towards Cheap, Flexible Robots”, Proceedings of the International Conference on Recent Advances in Mechatronics ICRAM 1995.

[8] Medhat K. Bahr Khalil,” Estimated versus calculated Viscous Friction Coefficient in Spool Valve Modeling”, Technical Conference on IFPE, 2008.

[9] Coulomb Friction (http://www.wisegeek.com/what-is-coulomb-friction.html)

[10] Tribology (http://en.wikipedia.org/wiki/Tribology)

[11] The Coefficient of Friction (http://www.mathsrevision.net/alevel/pages.php?page=79)

[12] Moment of inertia (http://en.wikipedia.org/wiki/Moment_of_inertia)

[13] Juan-Antonio Fernandez-Madrigal, Cipriano Galindo, Javier Gonza´ lez, Elena Cruz-Martin and Ana Cruz-Martın, "A software engineering approach for the development of heterogeneous robotic applications" ,Robotics and Computer-Integrated Manufacturing,Vol.24,pp.150-166,2008

[14] Sungho Jo, "Adaptive biomimetic control of robot arm motions”, Neurocomputing, Vol.71, pp.3625-3630, 2008

[15] Abdullah Ozer and S. Eren Semercigil, "An event-based vibration control for a two-link flexible robotic arm: Numerical and experimental observations", Journal of Sound and Vibration, Vol.313, pp.375-394, 2008

[16] Toshio Tsuji and Yoshiyuki Tanaka, "Bio-mimetic impedance control of robotic manipulator for dynamic contact tasks”, Robotics and Autonomous Systems, Vol.56, pp.306-316, 2008

[17] C.T. Freeman, Hughes, Burridge, Chappell, Lewin and Rogers, "A robotic workstation for stroke rehabilitation of the upper extremity using FES”, Medical Engineering & Physics, Vol.31, pp.364-373, 2009

[18] Yunquan Sun, DavidJ. Giblin and Kazem Kazerounian, "Accurate robotic belt grinding of work pieces with complex geometries using relative calibration techniques”, Robotics and Computer-Integrated Manufacturing, Vol.25, pp.204-210, 2009

[19] Nicola Vitiello , Tommaso Lenzi, Stefano Marco Maria De Rossi, Stefano Roccella and Maria Chiara Carrozza, "A sensor less torque control for Antagonistic Driven ,Compliant Joints", Mechatronics,Vol.20,pp.355-36,2010

[20] Celso De LaCruz , Teodiano Freire Bastos , Ricardo Carelli, "Adaptive motion control law of a robotic wheel chair" ,Control Engineering Practice,Vol.19,pp.113-125,2011

[21] C.M. Wronka and M.W. Dunnigan, "Derivation and analysis of a dynamic model of a robotic manipulator on a moving base”, Robotics and Autonomous Systems, Vol.59, pp.758-769, 2011

[22] Ashwini Shukla and Aude Billard, "Coupled dynamical system based arm–hand grasping model for learning fast adaptation strategies”, Robotics and Autonomous Systems, Vol.60, pp.424-440, 2012

[23] Hassan Azarkish, Said Farahat, and S.Masoud H. Sarvari, "Comparing the Performance of the Particle Swarm Optimization and the Genetic Algorithm on the Geometry Design of Longitudinal Fin ",World Academy of Science, Engineering and Technology,Vol.61,pp.836-839,2012

[24] [24] K Shivaprakash Reddy,PVK Perumal,B. Durgaprasad and M.A Murtaza, "Optimizing the Performance Evaluation of Robotic Arms with the Aid of Genetic Algorithm" , International Journal of Computer Applications, Vol. 51, No. 5 , pp.24-30 , 2012

[25] Erik Wernholt and Svante Gunnarsson, "Nonlinear Identification of a Physically Parameterized Robot Model", LiTH-ISY-R-2739, pp.1-6, Aug 2006

Method for Water Vapor Profile Retievals by Means of Minimizing Difference Between Estimated and Actual Brightness Temperatures Derived from AIRS data and Radiative Transfer Model

Kohei Arai [1]
(Graduate School of Science and Engineering)
Saga University
Saga City, Japan

***Abstract*— Method for water vapor profile retrievals by means of minimizing difference between estimated and actual brightness temperatures derived from AIRS data and radiative transfer model is proposed. Initial value is determined by linearized radiative transfer equation. It is found that this initial value determination method makes improvement of estimation accuracy together with reducing convergence time.**

Keywords-infrared sounder; non-linear optimization method; linearized inversion.

I. Introduction

Air-temperature and water vapor profiles are used to be estimated with Infrared Sounder data [1]. One of the problems on retrieving vertical profiles is its retrieving accuracy. In particular, estimation accuracy of air-temperature and water vapor at tropopause altitude is not good enough because there are gradient changes of air-temperature and water vapor profile in the tropopause so that observed radiance at the specific channels are not changed for the altitude.

In order to estimate air-temperature and water vapor, least square based method is typically used. In the process, Root Mean Square: RMS difference between observed radiance and calculated radiance with the designated physical parameters are minimized. Then the designated physical parameters including air-temperature and water vapor at the minimum RMS difference are solutions.

Typically, Newton-Raphson method which gives one of local minima is used for minimization of RMS difference. Newton-Raphson needs first and second order derivatives, Jacobean and Hessian at around the current solution. It is not easy to formularize these derivatives analytically. The proposed method is based on Levenberg Marquardt: LM of non-linear least square method. It uses numerically calculated first and second order derivatives instead of analytical based derivatives. Namely, these derivatives can be calculated with radiative transfer model based radiance calculations. At around the current solution in the solution space, directional derivatives are calculated with the radiative transfer model.

The proposed method is validated for air-temperature and water vapor profile retrievals with Infrared: IR sounder data derived from Atmospheric Infrared Sounder:/AIRS onboard AQUA satellite [2]-[7]. A comparison of retrieving accuracy between Newton-Raphson method and the proposed method based on LM method [8] is made in order to demonstrate an effectiveness of the proposed method in terms of estimation accuracy in particular for the altitude of tropopause [9]. Global Data Assimilation System: GDAS data of assimilation model derived 1 degree mesh data is used as truth data of air-temperature and water vapor profiles. The experimental data show that the proposed method is superior to the conventional Newton-Raphson method.

The following section describes proposed method for water vapor profile retrievals followed by experiments. Then finally, conclusion and some discussions are described.

II. Proposed Method

A. Radiative Transfer Equation

Radiative transfer equation is expressed with equation (1).

$$R\nu = (I_0)_\nu \tau_\nu(z_0) + \int_{z_0}^{\infty} B\nu\{T(z)\}K_\nu(z)dz \qquad (1)$$

where νdenotes wave number (cm-1), and

Rν: at sensor brightness temperature

(I0)ν : brightness temperature of ground surface

τ ν (z0): total column atmospheric transmittance

B{T(z)}ν: Planckian function of air temperature at the altitude of z

Kν(z): atmospheric transmittance at the altitude of z

This equation (1) can be linearized as follows,

$$R = BK \qquad (2)$$

where the number of unknown variables and the number of given equations are same. Therefore, it can be solved relatively easily. This solution from linear inversion provides initial value of the steepest descent method. Without this initial value, steepest descent method falls in one of local minima easily.

B. *Water Vapor Profile Retrival Method*

For instance, it can be solved based on steepest descent method as shown in equation (3)

$$R - R_0 = \frac{\partial R}{\partial q}\left(q - q_0\right) \qquad (3)$$

Also, it is possible to estimate water vapor profile to minimize the following covariance matrix of error,

$$\hat{x} = x_a + \left(A^T S_\varepsilon^{-1} A + S_a^{-1}\right)^{-1} A^T S_\varepsilon^{-1}\left(R - R_a\right) \quad (4)$$

where

x_a : Designated variable matrix

$\hat{x}$: Variable matrix for estimation

A: Jacobian Matrix

S_ε : Covariance matrix for measurement error

R : Observed brightness temperature

R_a : Estimated brightness temperature

Covariance matrix can be defined as equation (5).

$$S_{ij} = \varepsilon\left(x_i - \hat{x}_i\right)\left(x_j - \hat{x}_j\right)^T \qquad (5)$$

Jacobian Matrix can be expressed in equation (6).

$$A = \begin{pmatrix} \partial R_1/\partial q_1 & \partial R_1/\partial q_2 & \cdots & \partial R_1/\partial q_n \\ \partial R_2/\partial q_1 & \partial R_2/\partial q_2 & \cdots & \partial R_2/\partial q_n \\ \vdots & \vdots & \ddots & \vdots \\ \partial R_n/\partial q_1 & \partial R_n/\partial q_2 & \cdots & \partial R_n/\partial q_n \end{pmatrix} \quad (6)$$

C. *Steepest Descent Method (Non-linear optimization method)*

Steepest descent method can be represented in equation (7).

$$q_k = q_{k-1} + \alpha_k g_k \qquad (7)$$

where

q_k : estimated value at the iteration number k

g : updating vector

α : step width

Estimated value can be updated with the direction of g and with step size of α. Then estimation process is converged at one of local minima, not global optimum solution. This learning or updating process can be illustrated as shown in Figure 1. Initial value is derived from the linear inversion, K=B-1R.

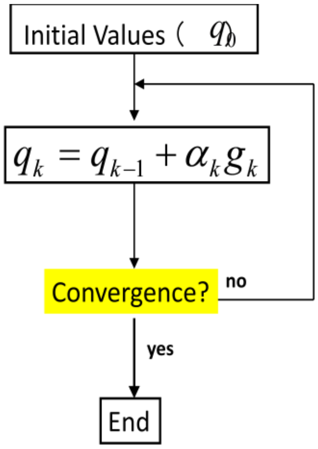


Figure 1. Process flow of steepest descent method

III. Experiments

A. *AIRS Channels Used for Estimation of Water Vapor Profile*

AIRS channels used for estimation of water vapor profile and the corresponding wave numbers are shown in Table 1.

TABLE I. AIRS Channels Used For Estimation Of Water Vapor Profile And The Corresponding Wave Numbers

Wave Number	1478	1483	1508	1514	1519	1541	1544	1558	1585
AIRS channel	1684	1692	1731	1740	1748	1761	1765	1786	1824

B. *AIRS Data Used*

AIRS data of Mexican gulf (Latitude: 20 degree North, Longitude: 92 degree West) which is acquired at 18:00 on November 16 in 2002 is used. Intensive study area is shown in Figure 2. Level 1B brightness temperature is shown in Figure 3. On the other hands, MODTRAN of radiative transfer code is used for estimation of at sensor brightness temperature. Mid. Latitude Winter of atmospheric model is assumed. Also carbon dioxide mixing ratio and total column water vapor as well as ozone is set at default values.

C. *Weighting Function and Brightness Temperature*

One of the examples of at sensor brightness temperature of AIRS (Water vapor absorption channels of the wave number ranges from 1460 to 1620 cm-1) is shown in Figure 4. Weighting function which corresponds to the water vapor absorption channels of wave number are shown in Figure 5.

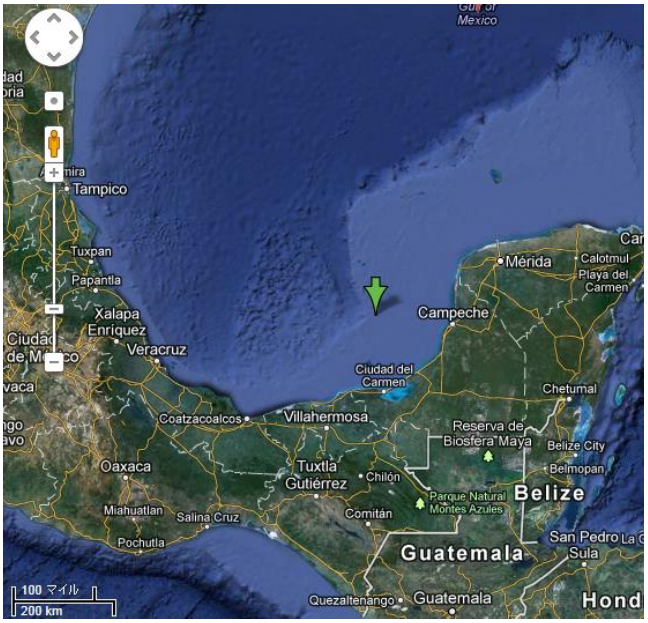


Figure 2. Intensive study areas (Mexican gulf)

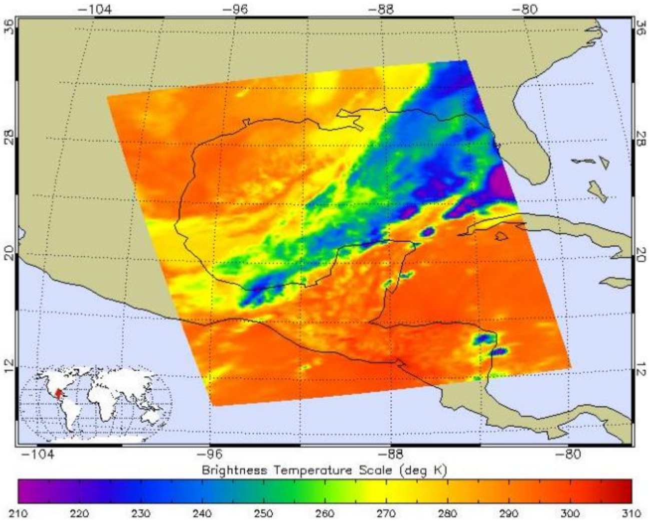


Figure 3. AIRS Level 1B of brightness temperature image

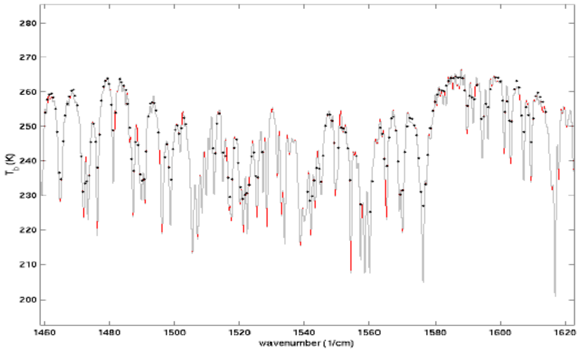


Figure 4. One of the examples of at sensor brightness temperature of AIRS (Water vapor absorption channels of the wave number ranges from 1460 to 1620 cm-1)

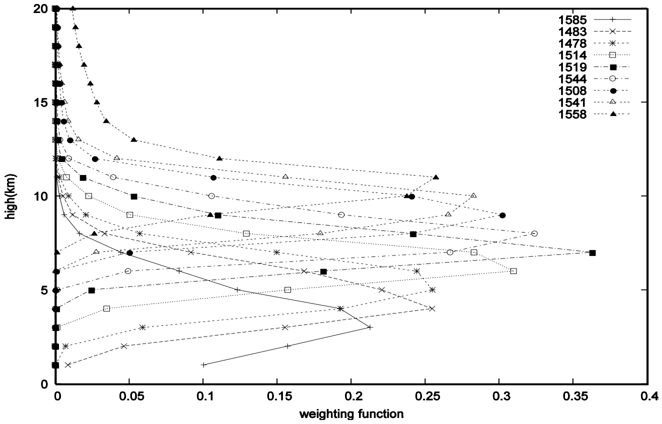
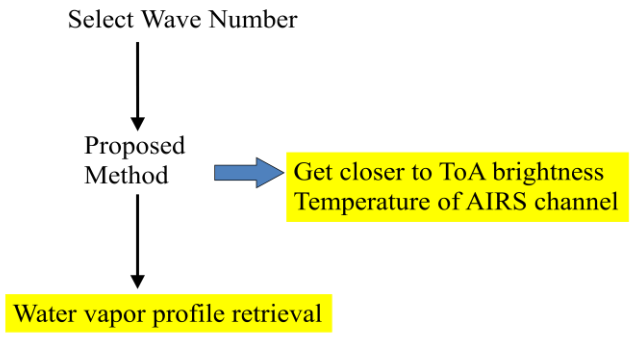


Figure 5. Weighting functions which correspond to the water vapor absorption channels of wave number

D. *Procedure for Water Vapor Profile Retrievals*

Based on the process flow which is shown in Figure 6, water vapor profile is retrieved. First, AIRS channels, wave numbers have to be determined. Then proposed method is applied to the AIRS data as well as MODTRAN derived at sensor brightness temperature.


Select Wave Number

Proposed Method

Get closer to ToA brightness Temperature of AIRS channel

Water vapor profile retrieval

Figure 6. Process flow for water vapor profile retrievals

In this process, initial value of the steepest descent method is given by the brightness temperature derived from linearized inversion. Minimizing the difference between estimated and actual AIRS brightness temperature, water vapor profile is retrieved. Figure 7 shows the difference of brightness temperature between the estimated Top of the Atmosphere: ToA brightness temperature for each AIRS channel of the corresponding wave number for retrieving water vapor profile and actual AIRS data derived brightness temperature at the convergence. The residual error is quite small Convergence condition is set at residual error in below 1x10-8.

E. *Retrieved Water Vapor Profile Based on the Proposed Method*

Retrieved water vapor profile by using the proposed method is shown in Figure 8. AIRS data used for the experiments is acquired on November 16 2002. Also, atmospheric model used for MODTRAN is Mid. Latitude Winter. Therefore, initial value of steepest descent method is close to the water vapor profile of Mid. Latitude Winter of atmospheric model.

Then the solution is updated to minimize the difference between AIRS derived brightness temperature and estimated ToA brightness temperature derived from MODTRAN.

Initial value is set by the calculated result from the linearized inversion, K=B-1R. The initial value is much closer to the default value of Mid. Latitude Winter model of MODTRAN. Therefore, it seems that the convergence stage of solution seems like a global optimum solution not one of local minima. If the initial value is set with random number, then convergence stage of solution looks like one of local minima, not global optimum solution. It took much longer time in this case while retrieval accuracy is not so good.

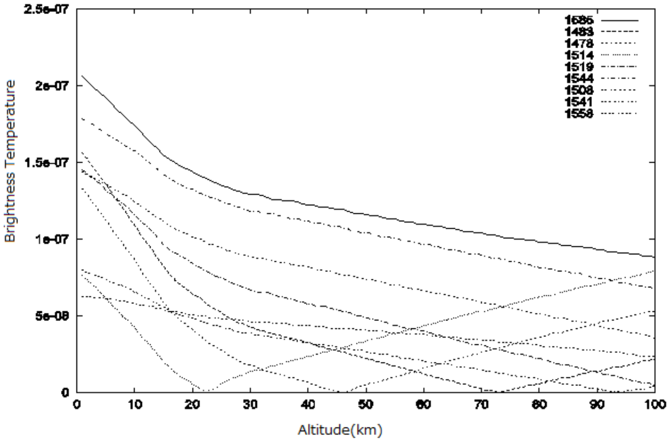


Figure 7. Examples of brightness temperature profile retrievals for the assigned AIRS channels, wave numbers

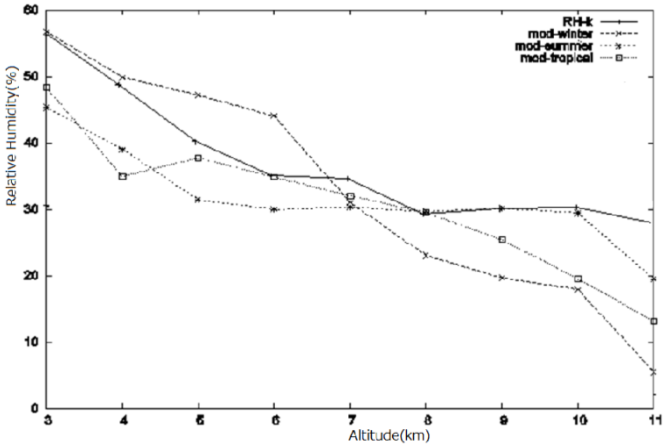


Figure 8. Retrieved relative humidity, water vapor profile and comparison to the default relative humidity profiles for atmospheric model of Mid. Latitude winter, and summer as well as tropic atmosphere

IV. CONCLUSION

Method for water vapor profile retrievals by means of minimizing difference between estimated and actual brightness temperatures derived from AIRS data and radiative transfer model is proposed. Initial value is determined by linearized radiative transfer equation.

It is found that this initial value determination method makes improvement of estimation accuracy together with reducing convergence time.

Initial value is set by the calculated result from the linearized inversion, K=B-1R. The initial value is much closer to the default value of Mid. Latitude Winter model of MODTRAN. Therefore, it seems that the convergence stage of solution seems like a global optimum solution not one of local minima. If the initial value is set with random number, then convergence stage of solution looks like one of local minima, not global optimum solution. It took much longer time in this case while retrieval accuracy is not so good

ACKNOWLEDGMENT

The author would like to thank Mr. Noriaki Yamada for his effort to conduct experiments.

REFERENCES

[1] Kohei Arai, Lecture Note on Remote Sensing, Morikita-Shuppan publishing Co. Ltd, 2004.

[2] NASA/JPL, "AIRS Overview". NASA/JPL. http://airs.jpl.nasa.gov/overview/overview/.

[3] NASA "Aqua and the A-Train". NASA. http://www.nasa.gov/mission_pages/aqua/.

[4] NASA/GSFC "NASA Goddard Earth Sciences Data and Information Services Center". NASA/GSFC. http://disc.gsfc.nasa.gov/AIRS/data_products.shtml.

[5] NASA/JPL "How AIRS Works". NASA/JPL. http://airs.jpl.nasa.gov/technology/how_AIRS_works.

[6] NASA/JPL "NASA/NOAA Announce Major Weather Forecasting Advancement". NASA/JPL. http://jpl.nasa.gov/news/news.cfm?release=2005-137.

[7] NASA/JPL "New NASA AIRS Data to Aid Weather, Climate Research". NASA/JPL. http://www.jpl.nasa.gov/news/features.cfm?feature=1424.

[8] Kohei Arai and Naohisa Nakamizo, Water vapor and air-temperature profile estimation with AIRS data based on Levenberg -Marquardt, Abstract of the 50th COSPAR(Committee on Space Research/ICSU) Congress, A 3.1-0086-08,995, Montreal, Canada, July, 2008

[9] Kohei Arai and XingMing Liang, sensitivity analysis for air temperature profile estimation method around the tropopause using simulated AQUA/AIRS data, Advances in Space Research, 43, 3, 845-851, 2009.

AUTHORS PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science, and Technology of the University of Tokyo from 1974 to 1978 also was with National Space Development Agency of Japan (current JAXA) from 1979 to 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He was appointed professor at Department of Information Science, Saga University in 1990. He was appointed councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was also appointed councilor of Saga University from 2002 and 2003 followed by an executive councilor of the Remote Sensing Society of Japan for 2003 to 2005. He is an adjunct professor of University of Arizona, USA since 1998. He also was appointed vice chairman of the Commission "A" of ICSU/COSPAR in 2008. He wrote 30 books and published 332 journal papers.

Peer Assignment Review Process for Collaborative E-learning: Is the Student Learning Process Changing?

Evelyn Kigozi Kahiigi
Department of Computer and Systems Sciences
Stockholm University, Sweden

Mikko Vesisenaho
University of Jyväskylä,
Finland

F.F Tusubira
Knowledge Consulting Ltd
Kampala, Uganda

Henrik Hansson
Department of Computer and Systems Sciences
Stockholm University, Sweden

Mats Danielson
Department of Computer and Systems Sciences
Stockholm University, Sweden

Abstract—**In recent years collaborative e-learning has been emphasized as a learning method that has facilitated knowledge construction and supported student learning. However some universities especially in developing country contexts are struggling to attain minimal educational benefits from its adoption and use. This paper investigates the application of a peer assignment review process for collaborative e-learning to third year undergraduate students. The study was aimed at evaluating the effect of the peer assignment review process on the student learning process. Data was collected using a survey questionnaire and analyzed using SPSS Version 16.0. While the student reported positive impact of the peer assignment review process in terms of facilitating students to put more effort and improve their work; quick feedback on their assignments; effective sharing and development of knowledge and information and the need of computer competence to manipulate the peer assignment review system, analysis of the quantitative data indicated that the process had limited effect on the learning process. This is attributed to lack of review skills, absence of lecturer scaffolding, low ICT literacy levels and change management.**

Keywords-Peer Review; Collaborative E-learning; Learning Process; Students; University; Uganda; Developing Country.

I. INTRODUCTION

The use of collaborative e-learning has been proposed to support learning by motivating students to share ideas, make critical judgment, reflect on their work as a result enforcing knowledge development and learning [1]. As a result the collaborative e-learning process has been widely recognized as a learning tool that supports students' performance [2]. Indeed, several learning theories have been used to explain the collaborative e-learning concept. Collaborative e-learning can be explained basing on social constructivism that relates to individuals constructing their knowledge through the process of negotiating meanings with others within the learning community [3]. With the social constructivism perspective the pedagogical approach is shifted from learning that focuses on delivery of content knowledge to collaborative learning aimed to facilitate acquisition of higher learning skills [3]. From a constructivist perspective of learning associated with Vygotsky's [4] zone of proximal development, the learner's level of understanding and cognitive development are attained through social interaction and collaboration. This implies that collaborative e-learning can allow learners develop an understanding and master various aspects in a course better than when working alone. Piaget's [5] work views collaborative e-learning as an active process that engages students in learning by giving them some ownership in their instruction. Piaget affirms that having access to peers work can provide new perspectives that challenge the student's understanding. Feedback received may cause cognitive dissonance to encourage students to modify his or her concepts resulting into new learning. By working together, students are able to accomplish and learn more than they could individually [6]. Students are given an opportunity to judge and compare their work with peers' work resulting in some level of understanding and knowledge creation. Indeed Topping et al.[7] view feedback as an integral part of a learning process through which students construct knowledge and develop their learning.

The collaborative e-learning concept using peer reviews has been investigated in numerous studies and learning scenarios that have facilitated knowledge construction and supported student learning. Sahin [8] validated peer evaluation in higher education and established a similarity with lecturer evaluation. Richardson et al.[9] evaluated the effectiveness of a peer feedback strategy in asynchronous online discussion. Students perceived the peer feedback as having impacted on their learning at a higher cognitive level, such as critical thinking skills. Lan et al. [10] on the other hand developed a conceptual framework for providing intelligent support through agent negotiation and fuzzy constraints to enhance the effectiveness of peer assessment. Student's performance significantly improved, the negotiation mechanism improved the assessment accuracy and thus students accepted the assessment results and reflected upon their work. Although benefits derived from peer review have been acknowledged and demonstrated, studies have also revealed challenges. Challenges reported in Kahiigi et al. [11] and Sahin [8] include students lack of skills to engage in meaningful reviews that are informative; validity and reliability of grades given by students resulting from lack of expertise and potential bias; negative attitude towards peer reviews; students being uncomfortable in carrying out assessments with a notion that it is the teachers' responsibility to assess and award grades, thus considering it an additional burden. These challenges have resulted into high levels of subjectivity thus limiting the adoption and use the peer review concept.

In this regard, the study presented in this paper aims to evaluate the effectiveness of the peer assignment review process for collaborative e-learning in an undergraduate Database Management Systems course. Specifically the study analyses the impact of the peer assignment review process and the change in the student learning process. The research questions discussed at this stage are: Does collaborative e-learning support student learning? And if so how has the learning process changed? In this paper collaborative e-learning is defined as a learning method that facilitates knowledge construction, negotiating meanings and solving problems to achieve a learning goal through mutual interaction between students using Information and communication technology (ICT).

II. Methodology

A. *Case Description*

The present study was carried out at Makerere University, College of Computing and Information Sciences between October and December 2011 with 998 third year undergraduate students enrolled in a Database Management Systems (DBMS) course. The DBMS course was a cross cutting course for students registered for Bachelors of Science in Software Engineering (BSE), Bachelors of Science in Computer Science (CSC), Bachelors of Information Technology (BIT) and Bachelors of Information Systems (BIS). This course was unique as it was taken by students from various disciplines and had a large student population. This DBMS course aimed to provide students with a strong foundation in systematic approaches to design and implementing of database applications and to provide a practical experience and knowledge in developing database driven applications in real world scenarios.

The DBMS course selection was based on: a) students using e-learning in their learning activities; and b) and the will of the lecturer to participate and drive students' involvement in the study. Noteworthy is that while all students registered to the different courses were part of the studies; involvement in the survey was voluntary.

B. *Case Study Procedure*

The DBMS course was traditionally taught by three lecturers following the same course outline with lectures held in a classroom setting on campus. The peer assignment review process was embedded within the course and marks earned on the assignment formed part of the final course evaluation. As a result student participation in the peer assignment review process was mandatory and participation in the study was voluntary. While the course had other course activities within an online and traditional learning environment, this study focused on the peer assignment review process that was introduced in the student's first course assignment.

Students followed the peer assignment review process stages these were: familiarization, assignment, review and feedback [11].During the familiarization stage students were introduced to the peer assignment review process and had a demonstration of the peer review application integrated into the Makerere University E-learning Environment (MUELE), the learning management system based on Moodle. A question and answer session was scheduled at the end of the demonstration to facilitate a deeper understanding and elaboration of peer assignment review process. The familiarization stage aimed to equip students with peer review skills. At the assignment stage, students were then given the assignment which was supposed to be submitted online using MUELE. After the submission deadline, which was scheduled to take place during the classroom sessions, the lecturers presented and discussed the marking criteria with the students. This exposed the students to possible answers and explanations that merited scoring.

The review stage, involves students being assigned two peer submissions anonymously. The students were required to download and review the submissions based on the set criteria, in addition to making constructive comments for each review. As a motivation, students were awarded 5 marks for each completed review. The feedback stage marked the end of the peer assignment review process. Students received feedback from their two peers and grades awarded by the lecturers. If students were not satisfied with the feedback received then they would flag the review which would then be moderated by the lecturer. Students sent emails to the course mailing list in case they needed help with using the application.

C. *Data Collection*

During the second half of the semester, a survey questionnaire was delivered. The questionnaire aimed to capture student's experience and willingness to adopt the peer assignment review process in their learning activities. In addition, the questionnaire aimed to elicit the students' understanding of the potential pedagogical benefits. The validity of the questionnaire was based on two aspects: First, the questionnaire items used were contextually modified from adapted and modified from De Raadt et al.[12] and Wood & Kurzel [13]. Secondly, the questionnaire was pre-tested on 10 randomly selected students in the DBMS class. The pre-test aimed to examine the general structure, clarity, and relevance of the questionnaire items. At completion of the pre-test, feedback received from the participants was used to modify the questionnaire.

D. *Data Analysis*

Data obtained from the survey was imported into SPSS for analysis. Ordinal logistic regression (OLR) was used to study the effect of independent variables on the dependent/outcome variable. Ferdousi & Levy [14] affirms that OLR does not require the assumption of linearity in the relation between independent and dependent variables. It estimates the magnitude of the effect of the independent variables on the dependent variable, thus making it superior in predicting the likelihood of dependent variable using independent variables [15]. Ordinal logistic regression is used when developing models to predict ordinal variables [16].

In the study, the learning process was treated as the dependent/outcome variable which constituted of five items (Ques_5A, Ques _5B, Ques _5C, Ques _5D and Ques _5E). The independent variables constituted of 18 items (Ques_10A - Ques 10D = 4 Items; Ques_11A - Ques_11F = 6 Items and Ques_12A - Ques_12H = 8 Items). The dependent and independent variables were scored on a likert scale (1-Strongly Disagree, 2-Disagree, 3-Neutral, 4-Agree, 5-Strongly Agree). The ordinal regression model was fitted to each of the five dependent/outcome variables of the learning process and analysed separately in order to observe the effect of the various independent variables concerning peer assignment review process, The study had a different hypothesis for each of the dependent variables. The independent variables were categorical thus treated as factors.

The purpose of this study was to establish the perceived students change in the learning process as a result of the peer assignment review process. As a result the analysis focused on two aspects: Firstly, establishing if the models improve the ability to predict the outcome. Secondly, ascertaining which variables (independent variables) related to the peer assignment review process has a significant effect on the learning process items (dependent variable).

III. Results

A. Descriptive Results

Out of the 998 students who were enrolled to the Database Management Systems course, 458 students voluntarily responded to the survey, of which 401 responses were usable. This accounted for 87.6% valid students responses with a gender composition of 42% female and 58% male distribution. 13.7% were registered BIS students, 36.4% were BIT students, 10.7% were BSE students and 39.2% were CSC students. The 18 questionnaire items related to the peer assignment review process were examined to ascertain the change in the participants learning process. The questionnaire items presented a high level of reliability with a Cronbach alpha coefficient of 0.912. George & Mallery (2003) indicate that a Cronbach alpha coefficient >.8 provides a good measure of internal consistency of items in the scale.

B. Examining the Change in the Learning Process

Results indicate that variables that had a significant effect on the student learning process items in relation to the peer assignment review process at 95% level of confidence. This implies that the significant variables led to changes in the learning process with everything held at a constant. The model fitting statistics indicated that the observed data was consistent with the estimated values of the fitting models as follows, for Ques_5A (χ^2 = 189.856, df =72 and sig = .000); Ques_5B (χ^2 = 137.365, df =72 and sig = .000); Ques_5C (χ^2 = 151.096, df =72 and sig = .000); Ques_5D (χ^2 = 168.814, df =72 and sig = .000); Ques_5E (χ^2 = 207.990, df =72 and sig = .000). The result indicates that the models are likely to predict the outcome since they are significant. In relation to the peer assignment review process variables and their effect on the learning process, parameter estimates for independent variables derived from each of the models were analysed (Table 1).

TABLE I. Parameter Estimates For Variables Related To The Peer Assignment Review Process On The Learning Process

Variables		Estimates	Sig.
Question 5A. The assignment inspired me to reflect on my use and understanding of course concepts			
[Ques_12B=2]	Peer assignment review process made me more interested in the topic	-2.026	.014
Question 5B. The assignment inspired me to search and learn beyond the material provided to me in class			
[Ques_11F=3]	Peer review feedback has added value for students	-.797	.027
[Ques_12C=3]	Peer assignment review process motivated me to improve my work	-1.579	.002
Question 5C. I received enough support from other students to complete the assignment (Group discussions)			
[Ques_10C=3]	Seeing other students using the system encouraged me to use it also	-1.029	.013
[Ques_11A=4]	I found the peer review process helped me to better reflect on my own work	-.844	.006
[Ques_11F=2]	Peer review feedback has added value for students	1.783	.017
[Ques_12D=1]	I felt secure about using the peer assignment review application	-3.020	.007
[Ques_12G=1]	I got assistance from fellow students when I failed to use the peer review system	-1.986	.000
Question 5D. When I saw other students assignments I compared them to my own assignment			
[Ques_10D=1]	Completing reviews anonymously allowed me to give feedback without bias	-2.525	.008
[Ques_11A=3]	I found the peer review process helped me to better reflect on my own work	-1.637	.005
[Ques_11C=3]	I was able to improve on my quality of assignment as a result of participating in the peer review process	-1.221	.008
[Ques_11D=3]	Feedback about my assignment came quickly from my peers than from the lecturer	-1.142	.005
[Ques_12E=4]	I liked reviewing other students assignments	.863	.012
[Ques_12G=3]	I got assistance from fellow students when I failed to use the peer review system	-1.691	.001
[Ques_12H=1]	I would be happy to use the same submission and review system in other courses	-1.804	.005
Question 5E. Through completing the reviews of other students work I developed a better understanding of the concepts covered in the assignment and the course			
[Ques_10B=2]	Communicating with other students through reviewing their assignments gave me the sense of belonging to the class	-1.954	028
[Ques_11E=2]	Peer review allows for effective sharing and development of knowledge and information	-2.313	023
[Ques_12F=1]	I was confident in carrying out the peer assignment review	-4.141	015

- Ques_5A. The assignment inspired me to reflect on my use and understanding of course concepts

Results indicate that Ques_12B=2 (peer assignment review process made students more interested in the topic) has a significant negative effect (Estimate = -2.026, p<.05) on Ques_5A (The assignment inspired me to reflect on my use and understanding of course concepts). This implies that for the participants who strongly agreed that the assignment inspired me to reflect on my use and understanding of course concepts, tended to strongly disagree that the peer assignment review process made students more interested in the topic.

- Ques_5B. The assignment inspired me to search and learn beyond the material provided to me in class

The OLR analysis indicate that Ques_11F=3 (peer review feedback has added value for students) and Ques_12C=3 (the peer assignment review process motivates students to improve my work) have a significant negative effect (Estimate = -.797, p<.05) and (Estimate = -1.579, p<.05) respectively on Ques_5B (the assignment inspired me to search and learn beyond the material provided to me in class). The result indicate that participants who strongly agreed that the assignment inspired them to search and learn beyond the material provided to me in class, tended to strongly disagree that peer review feedback has added value for students and that the peer assignment review process motivates students to improve my work.

- Ques_5C. I received enough support from other students to complete the assignment

From the OLR analysis, it was established that Ques_10C=3 (seeing other students using the system encouraged students to use it also), Ques_11A=4 (I found the peer review process helped students to better reflect on my own work), Ques_12D=1 (I felt secure about using the peer assignment) and Ques_12G=1(I got assistance from fellow students when I failed to use the peer review system) have a significant negative effect (Estimate = -1.029, p<.05), (Estimate = -.844, p<.05), (Estimate = -3.020, p<.05), (Estimate = -1.986, p<.05) respectively on Ques_5C (I received enough support from other students to complete the assignment),while Ques_11F=2 (peer review feedback has added value for students) has a positive significant effect (Estimate = 1.783, p<.05). These results indicate that participants who strongly agreed that they received enough support from other students to complete the assignment, tended to strongly disagree that seeing other students using the system encouraged students to use it also, the peer review process helped students to better reflect on my own work, they felt secure about using the peer assignment and that they got assistance from fellow students when I failed to use the peer review system. In addition results also imply that participants who strongly agreed that they received enough support from other students to complete the assignment, tended to strongly agree that peer review feedback has added value for students.

- Ques_ 5D. When I saw other students assignments I compared them to my own assignment

Results derived from the analysis in Table 1 indicated that Ques_10D=1 (Completing reviews anonymously allowed me to give feedback without bias); Ques_ 11A=3 (I found the peer review process helped me to better reflect on my own work), Ques_11C=3 (I was able to improve on my quality of assignment as a result of participating in the peer review process), Ques_11D=4 (Feedback about my assignment came quickly from my peers than from the lecturer), Ques_12G=3 (I got assistance from fellow students when I failed to use the peer review system) and Ques_12H=1 (I would be happy to use the same submission and review system in other courses) have a significant negative effect (Estimate = -2.525, p<.05), (Estimate = -1.637, p<.05), (Estimate = -1.221, p<.05), (Estimate = -1.142, p<.05), (Estimate = -1.691, p<.05) and (Estimate = -1.804, p<.05) respectively on Ques 5D (When I saw other students assignments I compared them to my own assignment), while Ques_12E=4 (I liked reviewing other students assignments) had a positive significant effect (Estimate = .863, p<.05). These results indicate that participants who strongly agreed that when I saw other students assignments I compared them to my own assignment tended to strongly disagree that completing reviews anonymously allowed me to give feedback without bias, they were able to improve on my quality of assignment as a result of participating in the peer review process, feedback about my assignment came quickly from my peers than from the lecturer, they got assistance from fellow students when they failed to use the peer review system and happy to use the same submission and review system in other courses. Results also implied that participants who strongly agreed that when saw other students assignments they compared them to their own assignment also strongly agreed that they liked reviewing other students assignments.

- Question 5E. Through completing the reviews of other students work I developed a better understanding of the concepts covered in the assignment and the course

OLR analysis results presented in Table 1, showed that Ques_10B=2 (Communicating with other students through reviewing their assignments gave me the sense of belonging to the class), Ques_11E=2 (Peer review allows for effective sharing and development of knowledge and information) and Ques_12F=1 (I was confident in carrying out the peer assignment review) have a significant negative effect with (Estimate = -1.954, p<.05), (Estimate = -2.313, p<.05) and (Estimate = -4.141, p<.05) respectively on Ques_5E (Through completing the reviews of other students work I developed a better understanding of the concepts covered in the assignment and the course). This implies that participants who strongly agreed that through completing the reviews of other students work I developed a better understanding of the concepts covered in the assignment and the course tended to strongly disagree that communicating with other students through reviewing their assignments gave them the sense of belonging to the class, that the peer review allows for effective sharing and development of knowledge and information and that they were confident in carrying out the peer assignment review.

IV. Discussion

The study reported in this paper aimed at studying the relationship between various variables related to peer assignment review process and their effect on the perceived change in the student learning process. It was envisaged that students would learn from each other and look at different perspectives presented by peers in order to improve the quality of their work and enhance their understanding of the course concepts [2] as such supporting their learning process. The results derived from the study indicate that the effect of the peer assignment review process on the students learning process was limited, as a result accounting to the insignificant perceived change in the students learning process.

Results indicated that although the assignment inspired students to reflect on the use and understanding of the course concepts, the peer assignment review process did not make then interested in the topic. This can be attributed to the fact that peer assignment review process was not aligned with the course objectives, thus affecting the learning outcome and students' expectations. Biggs [17] refers to the term "constructive alignment" whereby the desired learning outcomes are communicated to students, and learning activities and assessment tasks are coordinated to achieve these outcomes. Worth noting is that the peer assignment review process was a new concept that the students and lecturers were not used to and constructive alignment between the learning outcomes, assessment evidence and learning experiences was lacking.

In addition, results show that the assignment inspired students to search and learn beyond the material provided in class, and the peer review feedback did not have added value for the students. It was observed that although the feedback received from the students was timely in some cases it was unsatisfactory. 97% of the student assignments were moderated by the lecturers, based on the fact that the students were not satisfied with the reviews they received from their peers. Studies such as Sharpe & Benfield [18] and Ramsey [19] on collaborative and peer learning observed that it is difficult to engage students beyond interaction and information-sharing to constructive peer reviews. It was further noted that the peer assignment review process did not motivate students to improve their work. The educational culture of the research context and in most developing countries is lecturer-centred with lecturers as providers of information and students as receivers of information. In an effort to leapfrog students into the new collaborative e-learning dimension, there is a need for lecturer scaffolding to support and drive the learning process. The lecturer assumes a facilitator role encouraging focused learning and facilitating constructive interactions and reviews during the learning process.

As reported in Cassidy [20] students expressed concern regarding their ability and that of others to carry out the reviews. Willey & Gardner [21] view feedback as arguably the most important part because of its potential to affect future learning and student achievement. If feedback is not focused correctly (to inspire and motivate students to learn rather than circumvent their reflection and thinking) it may encourage dependent rather than independent learning. Fordyce & Mulcahey [22] established that students do not naturally take to the role of critic, attributing it to students shying away from commenting on their peers, their fear of alienating fellow students or their lack of the critical skills necessary to carry out the reviews. This point to the need to empower students and create opportunities among students to practice reviews in order to develop the required peer review skills. Walker [23] reports a change in students' attitude towards a positive perception of the peer review process resulting from these opportunities. Increasing students' familiarity with the peer review process and improving their skills can alleviate the perceived difficult sense of responsibility among them [20]. This implies that the introduction of the peer assignment review process should be gradually implemented to allow for effective and sustainable change in the learning process. From a technology acceptance perspective it can be inferred that a person using a technology or an application should find it free of effort [24]. In some instances technology and online environments can be frustrating, pointing to the lack of technology skills among the students [25], affecting the level of student involvement in the peer assignment review process. Zhu et al. [26] assert that computer competence is a significant predictor of students' achievement in online courses.

The findings indicate that effective sharing and development of knowledge and information through the peer assignment review process had a negative significant effect on the students' learning process. This finding contradicts findings reported in Richardson, et al. [9] that the peer review can foster an authentic learning environment in which students actively construct knowledge through reading, questioning ideas and reflecting on their own and peers' work. Wilson [27] adds that developing shared understanding among students is achieved through group consensus on knowledge, communicating and discussing different ideas and receiving feedback. As a result students learn by explaining their ideas to peers while participating in the process of inquiry. It is through this process that cognitive functions such as critical thinking increase, thus facilitating the learning process [9].

Developing a community provides motivation for learning, encourages engagement and reduces isolation [28]. Students interact to construct meaningful and worthwhile knowledge, an aspect that is crucial in any learning environment [29]. This confirms previous claims that student interaction can be related to deep learning, critical thinking, higher cognitive development and long-term knowledge retention [30]. It was observed that the pedagogical culture in the research context did not support such collaborative/interactive engagements, probably because the sense of community and connectedness resulting from the interaction in the peer assignment review process was not significant in providing a sense of belonging among students. The results indicated, however, that participants liked reviewing other students' work and comparing it with their own. This result may be related to the fact that students were curious to ascertain how other students had performed in the assignment. This creates competition among students which can lead to improved learning [31]. In addition, Pare & Joordens [32] affirm that reviewing peers' work encourages deep analysis of students' own work, resulting in improved quality of work.

V. CONCLUSION

This paper presents and discusses the application of a peer assignment review process for collaborative e-learning to learning activities of DBMS course taken by third year students at Makerere University. The aim was to determine the effect of peer assignment review process on the student learning process. The four peer assignment review process stages;

familiarization, assignment, review and feedback [11] were applied to the students learning activities during their semester period. Recent studied have claimed that the peer review process supports the learning process [1, 9]. However the results of this study indicate that peer assignment review process is a new concept yet to be comprehended and thus had limited impact on the students learning. This can be attributed to several factors limiting effective adoption and use of the peer assignment review process.

It was observed that while students were aware of the benefits and challenges derived from adopting and using the peer assignment review process; they were also reluctant to fully embrace as an approach to support their learning process. This can be attributed to the fact that students were not familiar to the learning approach which was implemented in a short time and they lacked the maturity to use the collaborative e-learning approach. Another possible explanation could be that student's expectation with regards to the peer assignment review process was based on their experience in the traditional learning environment. From a behaviourism perspective, learning is an observable change in behaviour that can be achieved by applying the concept of drill and practice [33]. In a sense students are accustomed to a certain form of learning after practicing it for a period of time and hence adapt to change. It should be noted that the students were not fully prepared for the peer assignment review process. An implication for further studies can be gradual implementation of the peer assignment review process to allow for effective and sustainable changes in the learning process.

Furthermore, students lacked the skills to review and critique their peers assignment. Indeed, Fordyce & Mulcahey [22] assert that students do not naturally take to the role of critic, attributing it students shying away from commenting on their peers, fear of alienating fellow students or that they do not have the critical skills necessary to carry out the reviews. However it is worth noting that continual engagement of students in the peer assignment review process can develop the review skills resulting into constructive feedback. The study observed that the varying levels of ICT literacy impacted on the adoption and use of the peer assignment review process. While some students were quick to submit and carryout the reviews others were struggling, thus sought help on using the system. This points to the relevance of ICT skills development in the study context to facilitate effective adoption and use of collaborative e-learning [34] to support the learning process.

Managing change is another factor that was apparent in the study context to have negatively impacted on students when using the peer assignment review process as part of their learning activity. Using student to pedagogically support each other's learning process through the peer assignment review process, puts students at the centre of the learning process. However this study has observed that it is difficult to engage students beyond interaction and information sharing to constructive peer reviews. This can be attributed to the traditional learning environment students are used to and the students expectations for lecturers involvements and provision of guidance. The educational culture in most developing countries is lecturer-centered with lecturers as providers of information and students as receivers of information. In an effort to leapfrog students into the new collaborative e-learning dimension, there is need for lecturer scaffolding to support and drive the learning process. The lecturer assumes a facilitator role encouraging focused learning and facilitating constructive interactions and reviews during the learning process.

ACKNOWLEDGMENT

Special thanks to Prof. Gebrenegus Ghilagaber, Dr. E.S.K Muwanga-Zake, Asrat Temesgen and J.B Asiimwe for helping with the data analysis.

REFERENCES

[1] E. Eryilmaz, N. Alrushiedat, and J. Van der Pol. The Effect of Anchoring Online Discussion on Collaboration and Cognitive Load. in The Fifteenth Americas Conference on Information Systems. 2009. San Francisco, California.

[2] L. Chung-Hsien, S. Graf, K.R. Lai, and Kinshuk, Enrichment of Peer Assessment with Agent Negotiation. IEEE Transactions on Learning Technologies, 2011, vol 4, no.1, pp. 35-46.

[3] H.J. So and T.A. Brush, Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. Computers & Education, 2008. vol.51, pp. 318-336.

[4] L. Vygotsky, Mind in Society: The development of Higher psychological process, Cambridge: Havard University Press, 1978.

[5] J. Piaget, The Psychology of Intelligence, New York: Harcourt and Brace, 1950.

[6] S. Turner, M.A. Pérez-Quiñones, and J. Chase, Exploring Peer Review in the Computer Science Classroom. Computing Research Repository., 2009.

[7] K.J. Topping, E.F. Smith, I. Swanson, and A. Elliot, Formative Peer Assessment of Academic Writing between Postgraduate Students. Assessment & Evaluation in Higher Education, 2000, vol 25, no.2, pp. 149-69.

[8] S. Sahin, An Application of Peer Assessment in Higher Education. The Turkish Online Journal of Educational Technology, 2008, vol 7, no.2, pp. 5-10.

[9] J.C. Richardson, P.A. Ertmer, J.D. Lehman, and T.J. Newby. Using peer feedback in online discussions to improve critical thinking. in Proceedings of the Annual Meeting of the Association for Educational Communications and Technology,Anaheim, CA, 2007.

[10] C.H. Lan, S. Graf, K.R. Lai, and Kinshuk, Enrichment of Peer Assessment with Agent Negotiation. IEEE Transaction on Learning Technologies, 201, vol 4, no.1, pp. 35-46.

[11] K.E. Kahiigi, M. Vesisenaho, H. Hansson, M. Danielson, and F.F. Tusubira, Modelling a Peer Assignment Review Process for Collaborative E-learning. Journal of Interactive Online Learning, 2012, vol.11, no.2, pp. 67-79.

[12] M. De Raadt, M. Toleman, and R. Watson. Electronic peer review: A large cohort teaching themselves. in In Proceedings of the 22nd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE'05), Brisbane, QUT, Brisbane, 2009.

[13] D. Wood and F. Kurzel. Engaging students in reflective practice through a process of formative peer review and peer assessment. in ATN Assessment Conference 2008: Engaging Students in Assessment. Adelaide, 2008.

[14] B. Ferdousi and Y. Levy, Development and Validation of a Model to Investigate the Impact of Individual Factors on Instructors' Intention to Use E-learning Systems. Interdisciplinary Journal of E-Learning and Learning Objects, 2010, vol.6.

[15] B.J. Ferdousi, A Study of Factors that Affect Instructors' Intention to Use E-Learning Systems in Two-Year Colleges, in Graduate School of Computer and Information Sciences,Nova Southeastern University, 2009.

[16] J.P. Hoffmann, Generalized linear models: An applied approach, Boston, MA: Pearson Education Inc, 2004.

[17] J.B. Biggs, Teaching for quality learning at universityBuckingham, UK: Open University Press, 1999.

[18] R. Sharpe and G. Benfield, The Student Experience of E-learning in Higher Education: A Review of the Literature. Brookes eJournal of Learning and Teaching, 2005, vol.1, no.3.

[19] C. Ramsey, Using Virtual Learning Environments to Facilitate New Learning Relationships. International Journal of Management Education, 2003, vol. 3, no.2, pp. 31-41.

[20] S. Cassidy, Developing employability skills: peer assessment in higher education. Education + Training, 2006, vol.48, no.7, pp. 508-517.

[21] K. Willey and A. Gardner, Investigating the capacity of self and peer assessment activities to engage students and promote learning. European Journal of Engineering Education, 2010, vol. 35, no.4, pp. 429-443.

[22] B. Fordyce and S. Mulcahey. Overcoming obstacles to student collaboration in peer review of written work Collaboration and Active Learningm Available from: http://iutconference.com/model.pdf, 2012

[23] A. Walker, British psychology students perception of group work and peer assessment. Psychology Learning and Teaching, 2001, vol. 1, no.1, pp. 28-36.

[24] F.D. Davis, R.P. Bagozzi, and P.R. Warshaw, User acceptance of computer technology: A comparison of two theoretical models. Management Science, 1989, vol. 35, no.8, pp. 982-1003.

[25] V. Cantoni, M. Cellario, and M. Porta, Perspectives and challenges in e-learning: towards natural interaction paradigms. Journal of Visual Languages & Computing, 2004, vol. 15,no.5, pp. 333-345.

[26] C. Zhu, M. Valcke, T. Schellens, and Y. Li, Chinese Students' Perceptions of a Collaborative E-Learning Environment and Factors Affecting Their Performance: Implementing a Flemish E-Learning Course in a Chinese Educational Context. Asia Pacific Education Review, 2009, vol. 10, no.2, pp. 225-235.

[27] G. Wilson, Online interaction impacts on learning: Teaching the teachers to teach. Australasian Journal of educational Technology, 2004, vol. 20, no. 1, pp. 33-48.

[28] C. Gray and K. Smyth, Collaboration Creation: Lessons Learned From Establishing an Online Professional Learning Community. The Electronic Journal of e-Learning Volume Issue, 2012, vol. 10, no.1, pp. 60-75.

[29] D.R. Garrison, Online community of inquiry review: Social, cognitive and teaching presence issues. Journal of Asynchronous Learning Networks, 2007, vol. 11, no.1, pp. 61-72.

[30] B. De Wever, T. Schellens, M. Valcke, and H. Van Keer, Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. Computers & Education, 2006, vol. 46, pp. 6-28.

[31] R. Takaoka, M. Shimokawa, and T. Okamoto, A Development of Game-Based Learning Environment to Activate Interaction among Learners. IEICE TRANSACTIONS on Information and Systems, 2012. vol. E95-D, no.4, pp. 911-920.

[32] D.E. Pare and S. Joordens, Peering into large lectures: examining peer and expert mark agreement using peerScholar, an online peer assessment tool. Journal of Computer Assisted Learning, 2008, vol. 24, no.6, pp. 526-540.

[33] F. Mödritscher, E-Learning Theories in Practice: A Comparison of three Methods. Journal of Universal Science and Technology of Learning, 2006, vol. 0, no.0, pp. 3-18.

[34] K.E. Kahiigi, H. Hansson, M. Danielson, F.F. Tusubira, and M. Vesisenaho. Collaborative E-learning in a Developing Country: A University Case Study in Uganda. in 10th European Conference on e-Learning ECEL-2011, Brighton Business School, University of Brighton, UK, 2011.

AUTHORS PROFILE

Evelyn Kigozi Kahiigi holds a Masters degree in Computer Science and is currently pursuing a PhD in Computer and Systems Sciences at Department of Computer and Systems Sciences, Stockholm University, Sweden. Her particular research interest is in collaborative e-learning, with specific focus on how it can be adopted and used effectively in a developing country context.

Mikko Vesisenaho is a research coordinator at the Human Technology Unit, Agora Center at the University of Jyväskylä, Finland. Dr. Vesisenaho’s multidisciplinary educational and academic background is from education and computer science. He has several years experience of working on East African ICT education research projects.

Dr. F. F. Tusubira is the current CEO of UbuntuNet Alliance for Research and Education Networking. He is actively involved in ICT policy and regulation formulation.

Henrik Hansson is Associate Professor and head of research in IT and learning at the Department of Computer and Systems Sciences, Stockholm University, Sweden. He is interested in the technology and learning

Mats Danielson is Professor in Computer and Systems Sciences at Stockholm University, Sweden. He is also the Vice Dean of the Social Science department. He is interested in computers in society, including decision support and analysis.

Agent Oriented Software Testing – Role Oriented approach

N.Sivakumar
Department of Computer Science and Engineering
Pondicherry Engineering College
Puducherry, India

K.Vivekanandan
Department of Computer Science and Engineering
Pondicherry Engineering College
Puducherry, India.

***Abstract*—Several Agent Oriented Software Engineering (AOSE) methodologies were proposed to build open, heterogeneous and complex internet based systems. AOSE methodologies offer different conceptual frameworks, notations and techniques, thereby provide a platform to make the system abstract, generalize, dynamic and autonomous. Lifecycle coverage is one of the important criteria for evaluating an AOSE methodology. Most of the existing AOSE methodologies focuses only on analysis, design, implementation and disregarded testing, stating that the testing can be done by extending the existing object-oriented testing techniques. Though objects and agents have some similarities, they both differ widely. Role is an important attribute of an agent that has a huge scope and support for the analysis, design and implementation of Multi-Agent System (MAS). The main objective of the paper is to extend the scope and support of role towards testing, thereby the vacancy for software testing perception in the AOSE series will be filled up. This paper presents an overview of role based testing based on the V-Model in order to add the next new component as of Agent-Oriented Software testing in the agent oriented development life cycle.**

Keywords-Agent oriented software enginerring; Multi-Agent System; Role oriented testing.

I. Introduction

A software development methodology refers to the framework that is used to structure, plan, and control the process of developing a software system. A wide variety of such frameworks have evolved over the years, each with its own recognized strengths and weaknesses. Now an increasing number of problems in industrial, commercial, medical, networking and educational application domains are being solved by agent-based solutions [1]. The key abstraction in these solutions is the agent. An “agent” is an autonomous, flexible and social system that interacts with its environment in order to satisfy its design agenda. In some cases, two or more agents should interact with each other in a multi agent system (MAS) to solve a problem that they cannot handle alone. The agent oriented methodologies provide us a platform for making system abstract, generalize, dynamic and autonomous. This important factor calls for an investigation of suitable agent-oriented engineering frameworks and testing techniques, to provide high-quality software development process and products

Agent Oriented Software Engineering (AOSE) is an umbrella term in which several researches have been proposed on new varieties of metaphors, formal modelling approaches and techniques, and development methodologies and tools, specifically suited towards agent-oriented paradigm. Several AOSE methodologies [4] were proposed for developing software, equipped with distinct concepts and modelling tools, in which the key abstraction used in its concepts is that of an agent. Some of the more popular AOSE methodologies were MASCommonKADS (1996-1998), MaSE(1999), GAIA(2000), MESSAGE(2001), TROPOS(2002), PROMETHEUS(2002), ADLEFE(2002), INGENIAS(2002), PASSI(2002), AOR Modeling(2003).

Several AOSE methodologies were analysed and compared and found that the strong weakness observed from almost all the methodologies were, there is no proper testing mechanism for testing the agent-oriented software. Our survey states that the agent based software are currently been tested by using Object-Oriented (OO) testing techniques, upon mapping of Agent-Oriented (AO) abstractions into OO constructs[3]. However agent properties such as Autonomy, Proactivity, and Reactivity etc., cannot be mapped into OO constructs. There arises the need for proper testing techniques for agent based software.

Role is an important attribute of an agent that has a huge scope and support for the analysis, design and implementation of Multi-Agent System (MAS) [2]. A role can be defined as the capability enabler that exposes to the agent that plays it a set of actions [9]. Roles are created to do something and it has the responsibility of achieving specific system goals and subgoals. Roles provide a well-defined interface between agents and cooperative processes [10]. This allows an agent to read and follow, normative rules established by the cooperation process even if not previously known by the agent. The major motivation to introduce such roles is to increase the agent system’s adaptability to structural changes.

The main objective of our paper is to propose a testing mechanism to test a multi-agent system based on agent’s important mental state, the role.

II. Background And Related Work

A. Lifecycle Coverage of AOSE Methodologies

For designing and building software systems, several software development paradigms were proposed such as structural, procedural, declarative and object-oriented technique. Recently agent oriented paradigm is used for designing and building software systems which is considered to be an extension of object oriented paradigm. Agents are the software program which works on behalf of human to carry some task which has been delegated to it and take their own decision according to the requirement. Agent based systems are meant for solving complex problem that too in distributed environment.

A methodology is indented to provide guidelines at every stage of software development process such as requirement analysis, design, implementation and testing. Several AOSE methodologies were proposed in the literature and every methodology deals with analysis and design phase of the agent based software development and very little attention is made for implementation and testing. When compared to implementation, testing has been neglected overall by the methodologies. Based on the survey made on several AOSE methodologies, it is very clear that the existing AOSE methodologies does not support testing phase, stating that testing the agent system has been accommodated using the traditional and object-oriented testing techniques. Table.1 tabulates the AOSE methodologies and their corresponding lifecycle coverage [4].

TABLE I. L IFE CYCLE COVERAGE OF AOSE METHODOLOGIES

Sl.No	Name of the Methodology	Life-Cycle Coverage
1.	MAS-CommonKADS (1996)	Analysis Design
2.	MASE(1999)	Analysis Design
3.	GAIA (2000)	Analysis Design
4.	MESSAGE(2000)	Analysis Design
5.	TROPOS(2001)	Analysis Design Implementation
6.	PROMETHEUS(2002)	Analysis Design
7.	PASSI(2002)	Analysis Design Implementation
8.	ADELFE(2002)	Analysis Design Implementation
9.	INGENIAS(2002)	Analysis Design Implementation
10.	RAP(2003)	Analysis Design

B. *Testing in AOSE Methodologies*

The goal oriented testing methodology [5] contributes to the existing Tropos methodology by providing a testing process model, which complements the Tropos methodology and strengthens the mutual relationship between goals and test cases. The support for testing in prometheus methodology is limited to only debugging support [5].

Role is an important mental attribute of an agent and often agent changes its roles to achieve its designated goal. Roles are intuitively used to analyze agent systems, model social activities and construct coherent and robust teams of agents. Roles are a useful concept in assisting designers and developers with the need for interactions. Roles provide a well-defined interface between agents and cooperative processes. Their major motivation to introduce such roles is to increase the agent system’s adaptability to structural changes. Among the existing AOSE methodologies, Generic Architecture for Information Access (GAIA) methodology [6] and Multiagent Systems Engineering (MaSE) methodology [7] were role-based methodologies for development of multi-agent systems. The GAIA methodology models both the social aspect and agent internals aspect of MAS. The methodology covers the analysis and design phase. Role model and Interaction model are constructed in analysis phase. With reference to the analysis phase, the agent model, service model and acquaintance model are constructed in the design phase. During the analysis phase of MaSE[7], a set of roles are produced, that describes entities which perform some function within the system. Each role is responsible for achieving the goal.

III. PROPOSED WORK

A software is been tested at different abstraction level such as unit, integration, system and acceptance. The main objective of this paper is to propose an effective agent-oriented testing technique that suits specifically for an agent based system. The proposed testing technique is oriented towards role, which is one of the important state/attribute of an agent. Role is defined as a capability enabler that exposes to the agent that plays it a set of actions. The V-model [8] represented in figure 1 shows the development and its corresponding role based testing activity. The left branch of the V represents the specification flow, and the right branch represents the role oriented testing flow.

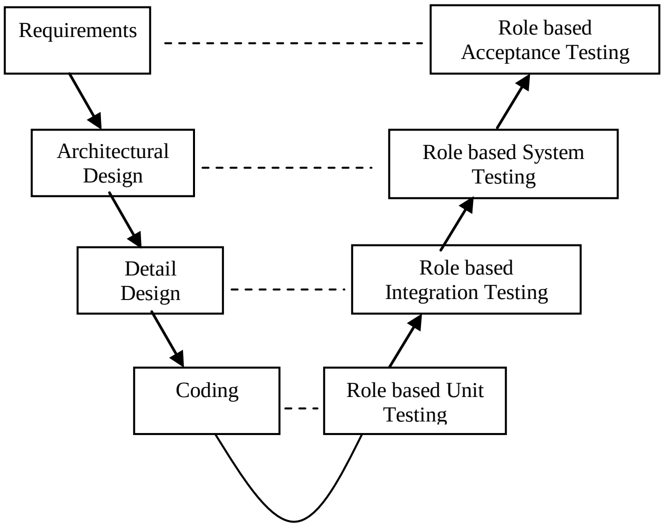

Requirements
Role based Acceptance Testing
Architectural Design
Role based System Testing
Detail Design
Role based Integration Testing
Coding
Role based Unit Testing

Figure 2. V-Model representation

A. *Role Model*

An attractive feature of agent-orientation is that it provides a powerful metaphor for describing, understanding and modeling information systems. A role is intended to enable software engineers to use it as a metaphor effectively to develop such cooperative information systems. The entire Role model is represented as Goal, Role and Responsibilities (GRRe).

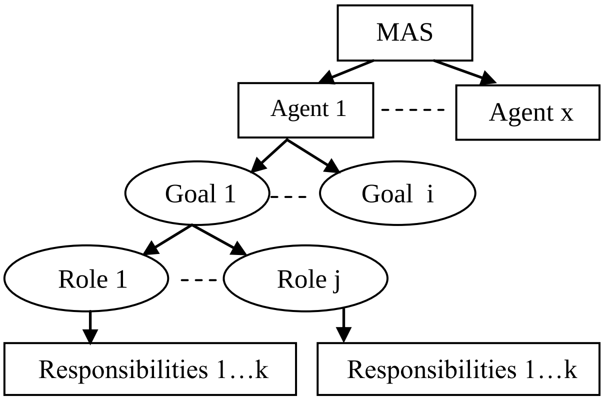


Figure 2. Role Model

A Multi-agent system is one which involves more than one agent that works in a coordinated way to achieve the overall goal of the system. Let us consider that, the number of agents involved in the MAS as 'x'. Every agent in the MAS is intended to achieve a goal in turn many sub-goals. Let us assume that the number of goals of an agent is 'i'.

The agent to achieve its goals and sub-goals has to play many roles within the agent as well as takes role in another agent i.e in certain cases the output from an agent might be the input to another agent. In such cases, the agent changes its current role and takes another role to accomplish the task given. Let us assume that the number of roles to achieve the agent's goal is 'j'. Moreover, a role is defined as the set of capabilities and expected behavior. In simple words, role is a set of actions or responsibilities to be handled by an agent. In this regard, a role can have 'k' number of responsibilities. Hierarchical representation of MAS down to responsibility is depicted in fig.3 and is mathematically been represented as below.

$$\sum_{x=1}^{m} A_n = \sum_{i=1}^{n} G_i \left(\sum_{j=1}^{o} R_j \left(\sum_{k=1}^{p} Re_k \right) \right)$$

Testing follows bottom-up approach in the proposed role-oriented testing. Test cases are written to test whether the responsibilities hold by the individual roles taken by the agent works as per the requirement.

Testing the responsibilities corresponding to their respective role ensures that the agent plays its role as per the intended goal. As long as the goals are getting satisfied, it is understood that the individual agent is performing well that suits the system. The following is the steps involved in role-oriented testing.

STEP 1: Select the Agent to be tested.

STEP 2: Identify Goals (G_i), Roles (R_j) and their corresponding Responsibilities(Re_k)

STEP 3: Design Role Model Diagram (A_x (G_i R_j Re_k))

STEP 4: Analyze role model (A_x(G_i R_j Re_k))

STEP 5: Define the interacting agents and situations.

STEP 6: Make the interacting agents as Pseudo agents.

STEP 7: Identify environmental factors pre-conditioning input trigger Re_k

STEP 8: Identify fulfillment criteria that satisfies Responsibility

STEP 9: Create test suite for Re_k corresponding to R_j

STEP 10: Run test cases

B. Role Oriented Unit Testing

Our testing approach focuses primarily on the smallest building block of the Multi-Agent System, the agent. The basic idea behind the unit testing in MAS is to verify whether the individual agent (unit) in isolation performs its responsibilities under various conditions.

Every individual agent has its own goal to be achieved and plans to do to fulfill the goal. In addition to goal and plan, role is one important mental state of the agent, which is defined as a set of capabilities and expected behavior. A role [9][10] can be represented as <Goal, Responsibilities, Protocol, Permissions>

- Goal, for which the agent playing this role is responsible
- Responsibilities, Which indicates the functionalities of agents playing such roles
- Protocol, which indicates how an agent playing such role can interact with agents playing other role
- Permissions, which are a set of rights associated with the role.

Let us consider an agent based university information system in which staff is one of the agent involved in the system. The goal of the staff agent is to be the best staff in the university. To achieve this goal, the staff agent has to take many roles such as teacher role, student counselor role, researcher role, administrator role etc., Every role has its own responsibilities, say for example, teacher role has the following responsibilities such as regular to class, handling classes properly, taking attendance, evaluating students, identifying weak students, etc.

Thus to test a single agent, scenarios (test cases) are to be developed to test whether all the responsibilities of the corresponding agent's role got satisfied. When all the roles are been tested and working fine, then by default we claim that the goal of the agent is been tested, thereby the individual agent is tested successfully. XML notations are used to define roles and their capabilities.

Figure.3 represents all the semantic information about the agent such as goal, role, responsibilities and protocols.

```
<Agent>
  <Goal>
    <Role1>
        <Responsibilities 1>
                <Protocol>
            .
        <Responsibilities n>
                <Protocols>
     </Role1>
     <Role n>
          <Responsibilities 1>
                <Protocol>
                .
          <Responsibilities n>
                 <Protocol>
       </Role n>
  </Goal>
 </Agent>
```

```
<Staff>
  <To be Best Staff>
    <Teacher>
         <Regular to class>
        <Handling classes>
        <Taking attendence>
        <Identifying weak students>
        <Evaluating students>
    </Teacher>

    <Researcher>
         <Involvement in research>
        <Publishing papers >
        <Usage of R&D activities>
        <Producing Ph.Ds>
     </Researcher>
  </To be Best Staff>
 </Staff>
```

Figure 3. XML Notation for defining Role and its

C. Role Oriented Integration Testing

Integration testing in conventional software development is that to progressively integrate the tested units (module, program, procedure, function) either incrementally or non-incrementally, so as to check whether the software units in an integrated mode are working properly. In an object-oriented software development, integrated testing is to verify the interaction among classes (interclass). The relationships among classes are the basic characteristics of an object-oriented system and define the nature of interaction among classes and objects at runtime. Multi-agent system is a logical collection of agents that interact with each other in a way that implements the functionality of the system. After ensuring that the individual agent in isolation is working as per the requirement, the next immediate step is to integrate the agents involved in the MAS so as to test the interaction, and communication among agents. Scenarios in which one agent interacts with another agent so as to comply with the role it holds are identified and tested. The protocol involved in communication among agent is also tested during integration.

D. Role Oriented System Testing

System testing verifies that all elements (hardware, people, databases) are integrated properly so as to ensure whether the overall product met its requirement and achieved the expected performance. System testing also deals with non-functional requirements of the software such as recovery testing, security testing, stress testing and performance testing. System testing in agent oriented approach will test the complete functionality and test the system as a whole. Here the perceptions and actions of all the agents are tested as a whole by providing proper test cases.

E. Role Schema

Role schema provides a well-defined interface between agents and cooperative processes. This allows an agent to read and follow, normative rules established by the cooperation process even if not previously known by the agent. Their major motivation to introduce such roles is to increase the agent system's adaptability to structural changes. Role schema involves role name, agent name, goal to be achieved, description of the role, protocol and related activities, permissions and responsibilities.

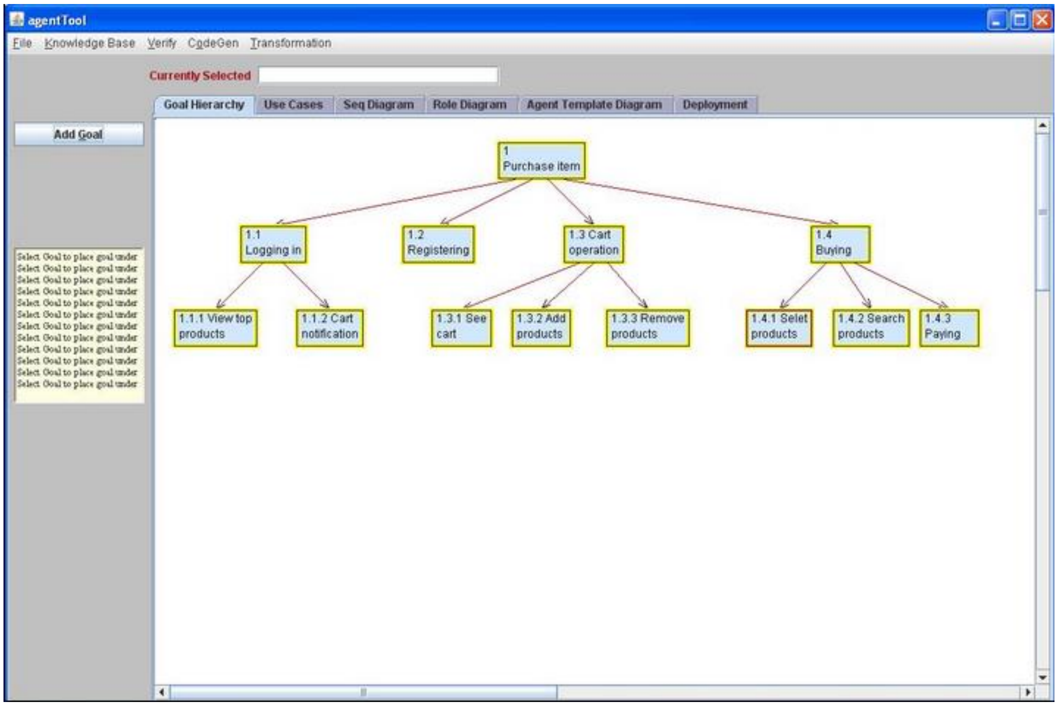
TABLE II. ROLE SCHEMA

Role Name : Teacher
Agent involved : Staff
Goal: To be the best staff in the university
Description: This role helps in identifying whether the teacher teaches well or not
Protocol and Activities: Be regular, Handling classes, Taking attendance, Evaluating students
Permissions: Read Request query, Result, Security policy**,** Change Result format // encrypt, Request format // decrypt
Responsibilities: *Activeness:* (Take attendance + Be regular+[Receive Questions from students+ Answer question from students]+ Teach subject+ Evaluate students + Submit result) *Completeness:* Lecturer is a good teacher

IV. CASE STUDY

To illustrate the role-based unit testing approach, an agent based online shopping system involving buyer agent, seller agent and bidder agent was developed using MaSE methodology. MASE is an iterative process. It deals the capturing the goals and refining the roles of an agent. It appears to have significant tool support. agentTool is a graphically based, fully interactive software engineering tool, which fully supports each step of MaSE analysis and design. Fig.3 shows the snapshot of the agentTool with which online shopping system is been analysed and designed. The analysis phase involves capturing goal, Applying use cases and Refining roles whereas the design phase involves Creating agent classes, Constructing conversations, Assembling agent classes and System design. Agent-based Online shopping system is been implemented using JADE(Java Agent Development

Framework), a software platform that provides basic middleware layer functionalities which are independent of the specific application and which simplify the realization of distributed applications that exploit the software agent abstraction.



Select Goal to place goal under
Select Goal to place goal under
Select Goal to place goal under
Select Goal to place goal under
Select Goal to place goal under
Select Goal to place goal under
Select Goal to place goal under
Select Goal to place goal under
Select Goal to place goal under
Select Goal to place goal under
Select Goal to place goal under
Select Goal to place goal under

Figure 4. Goal Hierarchy Diagram using agentTool

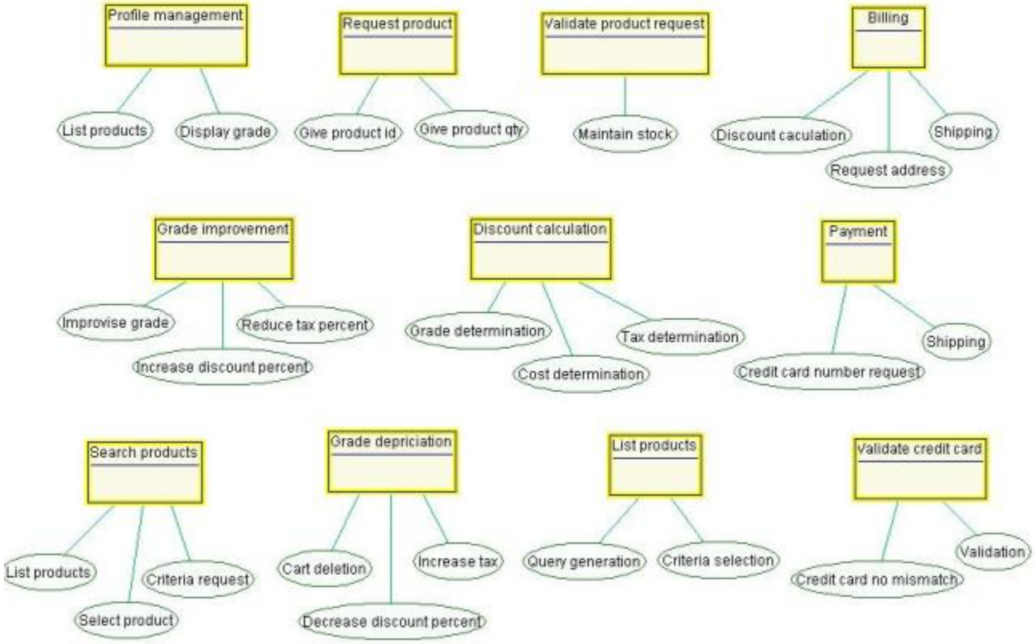
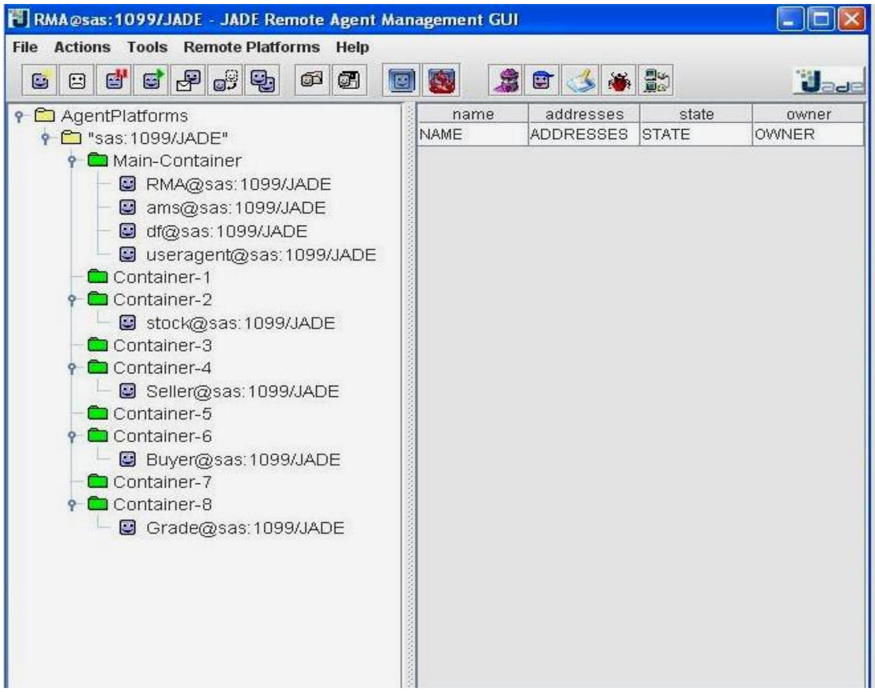


Figure 5. Role Diagram



- AgentPlatforms
  - "sas:1099/JADE"
    - Main-Container
      - RMA@sas:1099/JADE
      - ams@sas:1099/JADE
      - df@sas:1099/JADE
      - useragent@sas:1099/JADE
    - Container-1
    - Container-2
      - stock@sas:1099/JADE
    - Container-3
    - Container-4
      - Seller@sas:1099/JADE
    - Container-5
    - Container-6
      - Buyer@sas:1099/JADE
    - Container-7
    - Container-8
      - Grade@sas:1099/JADE

name	addresses	state	owner
NAME	ADDRESSES	STATE	OWNER

Figure 6. JADE Development Environment

V. Role Oriented Testing Of Online Shopping System

Testing is an important activity of the software development process. Efficiency of the testing techniques determines the quality of the software. AOSE methodologies struggle to compete with the existing programming paradigm as there is a lack of proper testing mechanism.

This paper comes out with a testing mechanism specifically for the agent software based on agent’s mental attribute, the role. The first step in the role-oriented testing is to identify the GRRe (Goal, Role, Responsibilities).

Let us consider a simple example derived from online shopping system. One of the goals of the online shopping system is to buy an item. To achieve this goal the agent has to take registrar role (registration process) and the Payer role (payment). Every role has its own responsibilities, say providing the registration detail is the responsibility of the registrar role and providing shipping detail is the responsibility of the payer role.

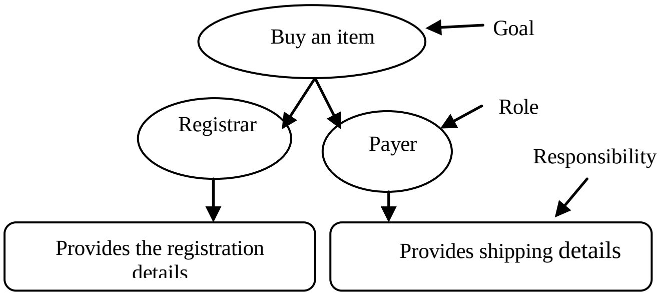


Figure 7. Sample Goal, Role, Responsibility representation

A. *Deriving Test cases*

According to our approach, the role of an agent comprises the logic of the test. As every role of an agent has number of responsibilities to get satisfied, the derivation of test case focuses on the responsibilities and thereby validates whether the role hold by the agent servers the purpose. Table 3 brings out the test case structure and some sample test case adopted for role oriented testing mechanism towards agent based online shopping system.

TABLE III. Test Case Structure For Role Based Testing

T.ID	TESTED AGENT	GOAL	ROLE	SITUATION	INPUT	EXPECTED RESULT	OBSERVED RESULT	RESULT
SE1	Seller Agent	Sell an item	Authenticator	Seller needs to authenticate elements	Username and Password	Authenticate and accept the Buyer	Buyer agent authenticated by seller	Passed

VI. Test Result Interpretation And Evaluation

In addition to the agent based online shopping system (P1) that we explored in the previous chapters we also developed few other agent based systems namely E-learning system (P2), Air ticket reservation system (P3), E-Novel System (P4) and E-Auction system (P5).

All the systems were developed using MaSE methodology. We intend to test all the agent based systems (P1 to P5) with the existing object oriented testing technique and with our proposed role based testing technique. We made a statistical comparison of object-oriented testing versus role-oriented testing techniques using the following metrics [11][12],

1. Defect Removal Efficiency (DRE),
   DRE = No. of defects resolved / Total no. of defects at the moment of measurement.

2. Test coverage (TC),
   TC= number of detected faults / number of predicted faults.

3. Test Case Effectiveness (TCE)
   No. of defects detected using test cases*100/Total no of defects detected

4. Number of Tests Per Unit Size (TPUS)
   Number of test cases per KLOC / FP where, KLOC represents Kilo Lines of Code & FP represents Function Point.

The above mentioned four metrics are calculated for five sample experimental projects P1 to P5. The calculated metrics are tabulated in Table.IV for further analysis and interpretation.

TABLE IV. Comparison Of Oo Testing Versus Role-Oriented Testing

Projects	DRE (%)		TC (%)		TCE		TPUS	
	OO	*RO*	*OO*	*RO*	*OO*	*RO*	*OO*	*RO*
P1	57.53	78.50	52.30	64.61	63.23	80.95	21.42	37.52
P2	43.54	82.67	45.67	68.23	64.32	82.33	25.89	35.07
P3	65.43	86.38	46.21	65.13	65.23	84.23	23.88	40.98
P4	53.76	82.34	45.52	69.32	64.12	82.45	22.67	37.61
P5	58.45	87.65	47.23	72.56	67.23	85.23	27.13	39.04

From the above table, it is very clear that our proposed role oriented testing is more efficient for testing the agent-based system rather than existing object-oriented testing technique. For further analysis and comparison the following graphs are plotted from the table data. From the graph, we may come to a conclusion that for testing an agent based system, an agent oriented testing technique say, role oriented testing will be more appropriate and effective too.

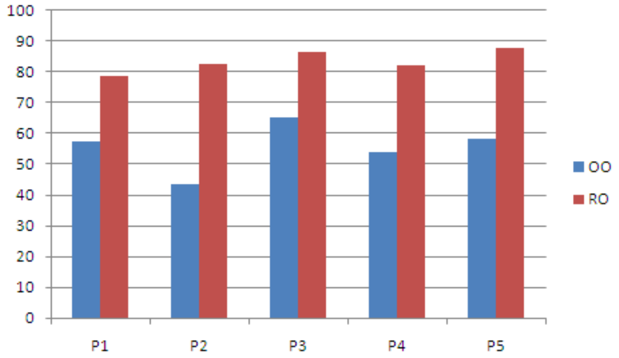


Figure 8. Defect Removal Efficiency

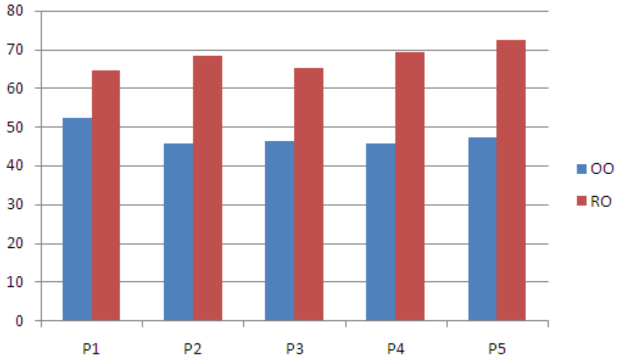
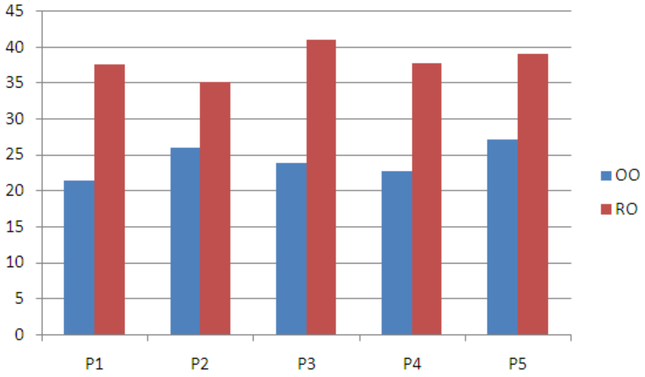


Figure 9. Test Coverage



A

Figure 10. Test Per Unit Size

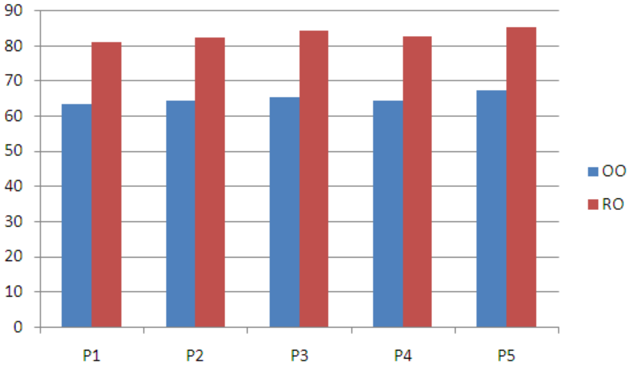


Figure 11. Test Case Effectiveness

VII. CONCLUSIONS

In this paper, a role oriented testing approach has been proposed for an agent system. The proposal helps the MAS developers in testing the individual unit of the agent based system namely, the agent. The testing mechanism is based on the role which is an important mental attribute of an agent. Every agent has its own role to perform so as to achieve its goal. Moreover the agents can even change their roles when the need arises so as to achieve the goal. Analysing the Goal-Role relationship, it is found that, as long as the agent performs its role properly, the goal of the system is been achieved by default. Thus testing whether the agent performs its role properly is a challenging task. This paved way for a role-oriented testing mechanism by which the role functionalities were tested by deriving appropriate test cases. To evaluate the proposed testing technique, 5 agent based systems were tested using object-oriented testing technique as well as role-oriented testing technique. The results of the testing techniques were tabulated and interpreted and it seems that the results are encouraging.

REFERENCES

[1] Shoham, Y. Agent oriented programming (Technical Report STAN-CS-90-1335) Stanford University: Computer science department, 1994.

[2] Giacomo Cabri, Letizia Leonardi, Luca Ferrari and Franco Zambonelli. Role-based Software agent interaction models: a survey. The Knowledge Engineering Review, Vol. 25:4, 397 419, 2010.

[3] Praveen Ranjan Srivastava , Karthik Anand V, Mayuri Rastogi, Vikrant Yadav, G.Raghurama. Extension of Object-oriented Software testing techniques to Agent Oriented software testing, in Journal of Object Technology, vol. 7, no. 8, November-December 2008, pp. 155-63.

[4] B. Henderson-Sellers and P. Giorgini. Agent-Oriented methodologies. Idea Group Inc., 2005.

[5] Duy Cu Nguyen, Anna Perini, Paolo Tonella. A goal-oriented software testing methodology. AOSE'07 Proceedings of the 8th international conference on Agent software engineering Springer –Verlag Berlin, Heidelberg-2008.

[6] B. Henderson, P. Giorgini. The Gaia Methodology for Agent-Oriented Analysis and Design Autonomous Agent and Multi-Agent Systems. 3, 285.312,2000 - 2000 Kluwer Academic Publishers.

[7] Wood, M. F. Multiagent system engineering: A methodology for analysis and design of muti-agent systems. Master thesis, School of Engineering, Air Force institute of technology, USA, (2000).

[8] Mailyn Moreno, Juan Pavon, Alejandro Rosete. Testing in Agent Oriented Methodologies. IWANN 2009, Part II, LNCS 5518, pp. 138–145, 2009.© Springer-Verlag Berlin Heidelberg 2009

[9] Haiping Xu, Xiaoqin Zhang, Rinkesh J. Patel. Developing Role-Based Open Multi-Agent Softwrae Systems. International Journal of Computational Theory and Practice Vol.2, No. 1, June 2007.

[10] Manjeet kumar. Roles and Ontology for Agent Systems. Global Journal of Computer Science and Technology Volume 11 Issue 23 Version 1.0 December 2011

[11] C. Wille, R. Dumke, and S, Stojanov. Quality Assurance in Agent-Based Systems Current State and Open Problems, Preprint No. 4. Fakultatfur Informatik, Otto-von-Guericke- Universitat, Magdeburg (2002).

[12] Daniel Galin. Software Quality Assurance. Pearson Education 2009.

An Advanced Certain Trust Model Using Fuzzy Logic and Probabilistic Logic theory

Kawser Wazed Nafi[1,4], Tonny Shekha Kar[1,4], Md. Amjad Hossain[2,4], M.M.A Hashem[3,4]
[1]Lecturer, Computer Science and Engineering, Stamford University Bangladesh,
[2]Assistant Professor, Computer Science and Engineering,
[3]Professor, Computer Science and Engineering,
[4]Khulna University of Engineering and Technology, Bangladesh

Abstract—**Trustworthiness especially for service oriented system is very important topic now a day in IT field of the whole world. Certain Trust Model depends on some certain values given by experts and developers. Here, main parameters for calculating trust are certainty and average rating. In this paper we have proposed an Extension of Certain Trust Model, mainly the representation portion based on probabilistic logic and fuzzy logic. This extended model can be applied in a system like cloud computing, internet, website, e-commerce, etc. to ensure trustworthiness of these platforms. The model uses the concept of fuzzy logic to add fuzziness with certainty and average rating to calculate the trustworthiness of a system more accurately. We have proposed two new parameters - trust T and behavioral probability P, which will help both the users and the developers of the system to understand its present condition easily. The linguistic variables are defined for both T and P and then these variables are implemented in our laboratory to verify the proposed trust model. We represent the trustworthiness of test system for two cases of evidence value using Fuzzy Associative Memory (FAM). We use inference rules and defuzzification method for verifying the model.**

Keywords-Certain trust; Certain Logic; Fuzzy Logic; Probabilistic Logic; FAM rule; Fuzzification; Defuzzification; Inference Rules.

I. Introduction

TRUST is a well-known concept in everyday life and often serves as a basis for making decisions in complex situations. There are numerous approaches for modeling trust concept in different research fields of computer science, e.g., virtual organizations, mobile and P2P networks, and E-Commerce. The sociologist Diego Gambetta has provided a definition, which is currently shared or at least adopted by many researchers. According to him "Trust is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action and in a context in which it affects its own action" [1].

The trustworthiness of the overall system depends on the following things:

a) The trustworthiness of the subsystems and atomic component independently from how these trust values are assessed.

b) Information on how the system combines its subsystems and components.

c) The knowledge about which subsystems and components are redundant.

Therefore, a major challenge of serving trust for the overall system is needed to consider that in real world applications the information about the trustworthiness of the subsystems and components itself is subject to uncertainty [1-4]. Besides, evaluating the models for the trustworthiness of complex systems are needed to be capable of modeling the uncertainty and also calculate and express the degree of uncertainty associated to the derived trustworthiness of the overall system. Trust is interrelated with people's everyday life. When people want to do some new things like selling or buying things from one to other, finding new materials from different source or some other things, concept of trust then arise. In the field of computer science and virtual world of modern technologies like virtual organizations, mobile or P2P networks and E-commerce [5-7], trust is very important. One to one conversion, data sharing is fully dependent on this trust on something. Different trust models have been developed now a day for serving this trustworthiness in the virtual communication world which is seen mostly worked on uncertainty.

Following the day by day improvements of the internet of services, the future internet based on Cloud computing IT systems will become highly distributed, dynamically composed and will be hosted and managed by multiple parties. But it is sorry to say at present people, enterprises, officials, organizations and corporate farms are still hesitating and feeling less of security and safety to move to the Cloud [8-10]. The reasons behind this are missing transparency, security concerns. So, both the users and providers and accreditation authorities are interested in evaluating the trustworthiness of a service, infrastructure or platform.

It is evident that the evaluation of the trustworthiness of complex systems is one of the major challenges in current IT research. Different trust models are now present in the world, which are dependent on uncertainty. [11-15] A new proposed model for solving this problem is Certain Trust Model (CTM) [1] which is used to calculate the trust of a system depending on recommendation of some experts, means on some certain values.

But, this model has some limitations. It has failed to apply fuzzy logic, probabilistic logic. This model is developed on the basis of human understanding. But for machine understanding and taking decisions, fuzzy logic is much helpful [16]. Again, it is much helpful for human beings to understand any situation and taking different type of decisions about something with the help of fuzzy logic [17-19] rather than other logics. The goal of our work is to extend representational model of CTM with the help of fuzzy logic, probabilistic logic so that the model can overcome its limitations. We have designed two new parameters for this purpose. These parameters are developed based on certain trust logic [2], which is dependent on CTM. The representational model of CTM will become friendlier to the users and the developers and make it more appropriate than the previous version of CTM.

Here, in this paper, section 2 describes the describes related work of our proposed work; section 3 describes briefly the model on which we work, section 4 describes our proposed model and implementation process, section 5 shows some case studies of our work, section 6 shows experimental results of our model which we had run in the lab.

II. RELATED WORK

Several number of ways are there for modeling (un-)certainty of trust values in the field of trust modeling in Cloud computing and internet based marketing sys-tem.[12, 13, 32] But, these models have less capability to derive trustworthiness of a system which are based on knowledge about its components and subsystems. The main challenge of these models is to find out good models for deriving trust using three ways:-

1. Trust from direct experience of a user,

2. Recommendations from third parties, ‘

3. Additional information.

These models help to save from robustness to attacks, e.g. misleading recommendations, Sybil attacks, etc. [20].

Fuzzy logic was used to provide trust in Cloud computing. Different types of attacks and trust models in service oriented systems, distributed system and so on are designed based on fuzzy logic system [21-24]. But it models different type of uncertainty known as linguistical uncertainty or fuzziness [17]. In paper [18], a very good model for E-commerce, which is based on fuzzy logic, is presented. But, this model also works with uncertain behavior. Belief theory such as Dempster-Shafer theory was used to provide trust in Cloud computing [20]. But the main drawback of this model is that the parameters for belief, disbelief and uncertainty are dependent on each other. It is possible to model uncertainty using Bayesian probabilities[25] which lead to probability density functions e.g., Beta probability density function. It is also possible to apply the propositional standard operators to probability density functions. But this leads to complex mathematical operations and multi-dimensional distributions which are also hard to interpret and to visualize. An enhanced model recently being developed for using in Cloud computing is known as Certain Trust. This model evaluates propositional logic terms that are subject to uncertainty using the propositional logic operators AND, OR and NOT[1-4].

III. CERTAIN TRUST MODEL

Certain Trust Model was constructed for modeling those probabilities, which are subject to uncertainty. This model was designed with a goal of evidence based trust model. Moreover, it has a graphical, intuitively interpretable interface [1] which helps the users to understand the model (Fig 1). The representational model focuses on two crucial issues

a) How trust can be derived from evidence considering context-dependent parameters?

b) How trust can be represented to software agents and human users?

For the first one, a relationship between trust and evidence is needed. For this, they had chosen a Bayesian approach. It is because it provides means for deriving a subjective probability from collected evidence and prior information [1]. At developing a representation of trust, it is necessary to consider to whom trust is represented. It is easy for a software component or a software agent to handle mathematical representations of trust. For it, Bayesian representation of trust is appropriate. The computational model of Certain Trust proposes a new approach for aggregating direct evidence and recommendations. In general, recommendations are collected to increase the amount of information available about the candidates in order to improve the estimate of their trustworthiness. This recommendation system needs to be integrated carefully for the candidates and for the users and owner of cloud servers. This is called robust integration of recommendations. In order to improve the estimate of the trustworthiness of the candidates, it is needed to develop recommendation system carefully. This is called robust integration of recommendations [1, 2].

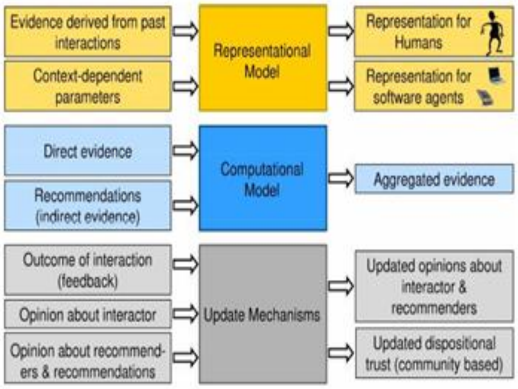

Evidence derived from past interactions
Context-dependent parameters
Representational Model
Representation for Humans
Representation for software agents
Direct evidence
Recommendations (indirect evidence)
Computational Model
Aggregated evidence
Outcome of interaction (feedback)
Opinion about interactor
Opinion about recommend-ers & recommendations
Update Mechanisms
Updated opinions about interactor & recommenders
Updated dispositional trust (community based)

Figure 1. Block diagram of Trust models

Three parameters used in certain logic: average rating t, certainty c, initial expectation f. The average rating t indicates the degree to which past observations support the truth of the proposition. The certainty c indicates the degree to which the average rating is assumed to be representative for the future. The initial expectation f expresses the assumption about the truth of a proposition in absence of evidence [1-4].

The equations for these parameters are given below:-

Equation for average rating, $$t = \begin{cases} 0.5 \; if \; r+s = 0 \\ r/(r+s) \quad else \end{cases} \qquad (1)$$

Here, r represents number of positive evidence and s represents number of negative evidence defined by the users or third person review system.

Equation for certainty,

$$c = \frac{N.(r+s)}{2.w.(N-(r+s)) + N.(r+s)} \qquad (2)$$

Here, w represents dispositional trust which influences how quickly the final trust value of an entity shifts from base trust value to the relative frequency of positive outcomes and N represents the maximum number of evidence for modeling trust. Using these parameters the expectation value of an opinion $E(t,c,f)$ can be defined as follows:

$$E(t,c,f) \;=\; t*c \;+\; (1-c)\;*\;f \qquad (3)$$

The parameters for an opinion o = (t, c, f) can be assessed in the following two ways: direct access and Indirect access. Certain Trust evaluates the logical operators of propositional logic that is AND, OR and NOT. In this model these operators are defined in a way that they are compliant with the evaluation of propositional logic terms in the standard probabilistic approach. However, when combining opinions, those operators will especially take care of the (un)certainty that is assigned to its input parameters and reflect this (un)certainty in the result. The definitions of the operators as defined in the CTM are given in the table 1.

TABLE I. DEFINITION OF OPERATORS

OR	$c_{A \vee B} = \quad c_A + c_B - c_A c_B - \frac{cA(1-cB)fB(1-tA)+(1-cA)cBfA(1-tB)}{fA+fB-fAfB}$ $t_{A \vee B} \;=\; \begin{cases} \frac{1}{cA \vee B}(c_A t_A + c_B t_B - c_A c_B t_A t_B) & \text{if } c_{A \vee B} \neq 0 \\ 0.5 & \text{else} \end{cases}$ $f_{A \vee B} \;=\; f_A + f_B - f_A f_B$
AND	$c_{A \wedge B} = c_A + c_B - c_A c_B - \frac{(1-cA)cB(1-fA)tB+cA(1-cB)(1-fB)tA}{1-fAfB}$ $t_{A \wedge B} = \begin{cases} \frac{1}{cA \wedge B}\left(c_A c_B t_A t_B + \frac{cA(1-cB)(1-fA)fBtA+(1-cA)cBfA(1-fB)tB}{1-fAfB}\right) & \text{if } c_{A \wedge B} \neq 0 \\ 0.5 & \text{else} \end{cases}$ $f_{A \wedge B} \;=\; f_A f_B$
NOT	$t_{\neg A} = 1 - t_A$, $c_{\neg A} = 1 - c_A$ and $f_{\neg A} = 1 - f_A$

IV. PROPOSED APPROACH AND USED CASES

For developing and exercising with our work, we have used a scenario from the field of cloud computing [1]. We have assumed that we are working to evaluate the trustworthiness of an organization or a simple office. We have worked mainly for the field of trade and business web pages. It has helped us to calculate the trust of the whole cloud computing system and also helped to make a system trustworthy to the user.

In this test system (figure 2), the server S directly relies on two subsystems or servers, S1 and S2. Subsystem S1 consist of two servers (A1 and A2), where at least one of the servers has to be available. Similarly, subsystem S2 is composed of two redundant data bases servers (only one need to be available).

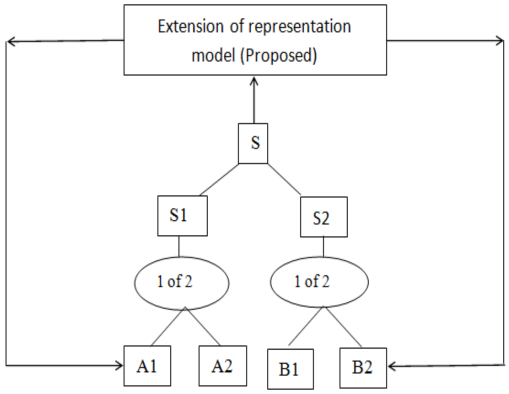


Figure 2. Assumed Cloud Architecture

Based on the description above and getting the information about the trust values of the atomic components, the evaluation of the trustworthiness of the complete system in the context of availability, can be carried out by evaluating the following propositional logic term:

$$(A1 \vee A2) \;\wedge\; (B1 \vee B2) \qquad (4)$$

According to the certain trust, the average rating t indicates the degree to which past observations support the truth of the proposition. t = 0 implies that there is only evidence contradicting the proposition. t = 1 implies that there is only evidence supporting the proposition. According to us, this average rating can be scaled according to the wish of the developer. E.g. One developer wants to scale the rating of his product from 1 to 5, means 1 is the lowest rate of the product and 5 is the highest rate. This is one type of ***Rating Based Trust Model*** [33] approach. So, for him, this average rating can be scaled between 1 and 5. From the rate of the product, one user or buyer can take his decision about the product and keep him safe from being betrayed. Again, another parameter, named certainty c, indicates the degree to which the average rating is assumed to be representative for the future. The higher the certainty, the higher is the influence of the average rating on the expectation value in relation to the initial expectation. When the maximum level of certainty is reached, the average rating is assumed to be representative for the future outcomes. c = 0 implies that there is no evidence available. c = 1 implies that the collected evidence is considered to be representative [1]. Here, the certainty c not only takes the value of 0 or 1 but also the values between 0 and 1. The scaling of this parameter depends on the interest of the developer or the manager of the office or industry. It mainly depends on the number of evidence. The last parameter of certain logic is initial expectation f. The initial expectation f expresses the assumption about the truth of a proposition in absence of evidence. It is helpful for the new owner or developer of a product to express his expectation about the product's service to humans. [1, 2, 3].

After taking the output from the CTM, we have applied *fuzzy logic* on it. It is an extension of the representational model of CTM. After getting value of c, t and f at the system top position S, our model starts working. We have plotted this in basic fuzzy logic system. After calculating, the result will be send to the lower level users, means to the lower level servers, PCs and system. According to our figure 2, these are A_1, A_2, B_1, B_2.

With the help of these parameters and operators derived from certain trust, we have defined two new parameters, Trust T and behavioral probability, P. Trust T is calculated from certainty c and average rating t. the equation is:

$$\text{Trust, T} = \frac{c*t}{Highscalingvalueofrating} * 100\% \qquad (5)$$

Here, High scaling value of rating means the upper value of the range of rating.

Calculating T, we have applied FAM rule of fuzzy logic for creating a relation between certainty c and average rating t. Trust T represents this relation in percentage such a way that the quality of the product can easily be understood.

Another parameter, behavioral probability, P, represents how the present behavior of the system varies from its initially expected value and it is proposed with the help of probabilistic logic [30].

It may be less, equal or higher than the initial expectation given by the system developer or the manager of the office. The equation for P is:

$$\text{Behavioral probability, P} = \frac{(T)-f}{f} * 100\% \qquad (6)$$

If T<f, lower probability to show expected behavior

If T>f, higher probability to show expected behavior

If T=f, balanced with the expected behavior

From the equation of P, values with two type magnitude have been found. If it is negative, then it is assumed that it will behave lower than the expected. If it is positive, then higher behavior will be shown by the system than the expected behavior of this system, which is defined by the developer or someone related to the system.

We can see following values for behavioral probability, P.

TABLE II. BEHAVIORAL PROBABILITY FOR DIFFERENT RANGES

Trust Ranges	Calculated P	Comment
1-20%	98%-60%	Lowest Behavior
21-40%	58%-20%	Lower Behavior
41-49%	18%-2%	Low Behavior
50%	0	Balanced
51-60%	2%-20%	High Behavior
61-80%	22%-60%	Higher Behavior
81-100%	62%-100%	Highest Behavior
>100%	100%	Highest Behavior

Where, initial expectation f is assumed to be 0.5; means showing 50% of accurate behavior of the system or the product. We have run this whole system in our lab and have got related results discussed above.

Observing the value of P, one can easily understand whether the system can fulfill his expectation or not according to the expectation of the developer about the whole cloud computing system or the trading product. These two parameters help both the developer and user.

A. Fuzzification and Defuzzification

A fuzzy logic system (FLS) can be defined as the nonlinear mapping of an input data set to a scalar output data [25-29]. A FLS consists of four main parts: fuzzier, rules, inference engine, and defuzzier. Fuzzy logic consists of following components are: -

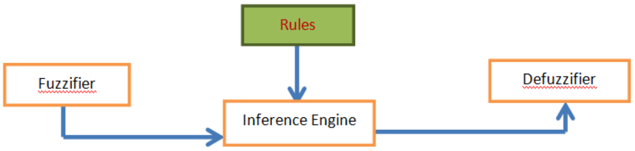


Figure 3. Basic components of fuzzy model

The process of fuzzy logic maintains the following steps: Firstly, a crisp set of input data are gathered and converted to a fuzzy set using fuzzy linguistic variables, fuzzy linguistic terms and membership functions. This step is known as fuzzification. Afterwards, an inference is made based on a set of rules. Lastly, the resulting fuzzy output is mapped to a crisp output using the membership functions, in the defuzzification step. Fuzzification is a process where inputs are a set of fuzzy inputs and the output is crisp values.

B. Fuzzy Inputs

According to Gaussian, the membership function for fuzzy input sets depends on two types of parameter, standard deviation **σ** and mean **c.** The equation for membership function is:-

$$f(x; \boldsymbol{\sigma}; \mathbf{c}) = \exp\left(\frac{-(x-c)^2}{2\sigma^2}\right) \qquad (7)$$

In designing fuzzy inference system, it is easy to understand that membership functions are associated with term sets, which normally appears in the antecedent or consequent of rules. We have divided parameter certainty c into five categories according to its values. They are:-

TABLE III. RANGES OF CERTAINTY

Class Name	Certainty Range Value	Symbols
Very Low	0.0-0.2	VLc
Low	0.1-0.4	Lc
Average	0.3-0.7	Avg.c
High	0.6-0.9	Hc
Very High	0.8-1.0	VHc

Following the same way, we have divided parameter average rating t into five categories according to its values in table IV:-

Though we classify the parameters value according to the ranges described above, it can be varied from persons to persons. For this, we take helps from fuzzy logic.

TABLE IV. RANGES OF AVERAGE RATING

Class Name	Avg. Rating Range Value	Symbols
Very Low	1.0-2.0	VLt
Low	1.5-3.0	Lt
Average	2.0-4.0	Avg.t
High	3.0-4.5	Ht
Very High	4.25-5.0	VHt

Following the Gaussian membership function equation, we have got the figures stated below:-

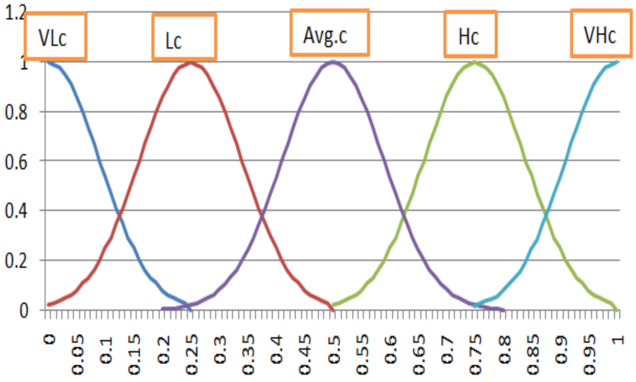


Figure 4. Membership Functions for certainty

Here X- axis represents the certainty deviation.

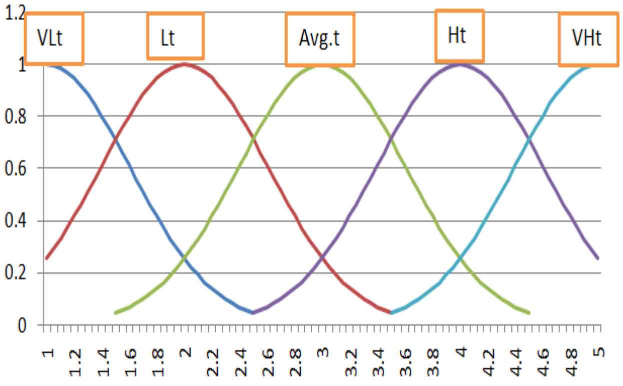


Figure 5. Membership Functions for average rating

C. Inference Rules

Fuzzy inference is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns discerned. There are two concepts of fuzzy logic systems [27]. They are: - linguistic variables and fuzzy if then else rule. The linguistic variables' values are words and sentences where if then else rule has two parts; antecedents and consequent parts which contain propositions of linguistic variables. Numerical values of inputs $x_i \epsilon U_i (i=1,2….n)$ are fuzzified into linguistic values, F_1, F_2…..F_n. Here F_i denotes the universe of discourse $U = U_1 * U_2 * …….. * U_n$. The output linguistic variables are $G_1, G_2, ……, G_n$. The if-then-else rule can be defined as: -

$R^{(j)}$: *IF* $x_i \epsilon F_1^{\,j}$ *and……and* $x_n \epsilon F_n^{\,j}$ THEN $y \epsilon G^j$. (8)

Where, j = 1,2,….., M. M is the number of rules.

According to rules discussed above, we have proceeded for our proposed model. There are 25 fuzzy rules in our extension model. They are (R represents rule):-

R1:- If certainty is very low and average rating is very low, then trust is very low.

R2:- If certainty is low and average rating is very low, then trust is very low.

R3:- If certainty is average and average rating is very low, then trust is very low.

R4:- If certainty is high and average rating is very low, then trust is very low.

R5:- If certainty is very high and average rating is very low, then trust is very low.

R6:- If certainty is very low and average rating is low, then trust is very low.

R7:- If certainty is low and average rating is low, then trust is low.

R8:- If certainty is average and average rating is low, then trust is low.

R9:- If certainty is high and average rating is low, then trust is average.

R10:- If certainty is very high and average rating is low, then trust is average.

R11:- If certainty is very low and average rating is average, then trust is very low.

R12:- If certainty is low and average rating is average, then trust is low.

R13:- If certainty is average and average rating is average, then trust is average.

R14:- If certainty is high and average rating is average, then trust is average.

R15:- If certainty is very high and average rating is average, then trust is high.

R16:- If certainty is very low and average rating is high, then trust is very low.

R17:- If certainty is low and average rating is high, then trust is low.

R18:- If certainty is average and average rating is high, then trust is average.

R19:- If certainty is high and average rating is high, then trust is high.

R20:- If certainty is very high and average rating is high, then trust is high.

R21:- If certainty is very low and average rating is very high, then trust is very low.

R22:- If certainty is low and average rating is very high, then trust is low.

R23:- If certainty is average and average rating is very high, then trust is average.

R24:- If certainty is high and average rating is very high, then trust is high.

R25:- If certainty is very high and average rating is very high, then trust is very high.

According to these inference rules stated above, we have got the fuzzy input sets shown in figure 4 and figure 5. From that, we have got figure 6 output crisp values.

D. *Fuzzy Outputs*

From the input fuzzy sets described above, passing those fuzzy sets through inference rules and fuzzy base rules, we get crisp values for our new parameter trust T. Plotting those values according to Gaussian membership function equation we have got the figure… for Trust T parameter. It can also be classified into five categories after finding out and plotting:-

TABLE V. RANGES OF OUTPUT TRUST

Class Name	Trust Range Value	Symbols
Very Low	0%-20%	VLT
Low	10%-40%	LT
Average	30%-70%	Avg.T
High	60-90%	HT
Very High	80%-100%	VHT

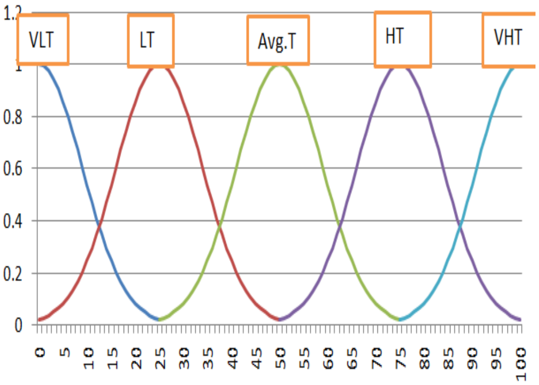


Figure 6. Membership Functions for Output Trust

Here X-axis represents the trust values.

E. *Defuzzification*

The input for the defuzzification process is a fuzzy set and the output of defuzzification process is a crisp value obtained by using some defuzzification method such as centroid, height and maximum. Among them, centroid defuzzification is used mostly. We use the following equation for applying Defuzzification method:-

$$y' = \frac{\sum_{i=1}^{n} yi\mu(yi)}{\sum_{i=1}^{n}\mu(yi)} \quad (9)$$

F. *Applying Implication Method*

The prerequisite for applying implication to any fuzzy set is finding out rule's weight.

We have found out the weights in figure 6 Membership functions output is de-fined as the weights for every rules.

The input for the implication process is a single number given by the antecedent, and the output is a fuzzy set. Implication is implemented for each rule.

Let, one of the rules is:-

"If certainty is high and average rating is aver-age, then trust is average."

Let, the value for certainty is C=0.7 and value for average rating is t = 3.0. Now, according to the implication method, we get the following output:-

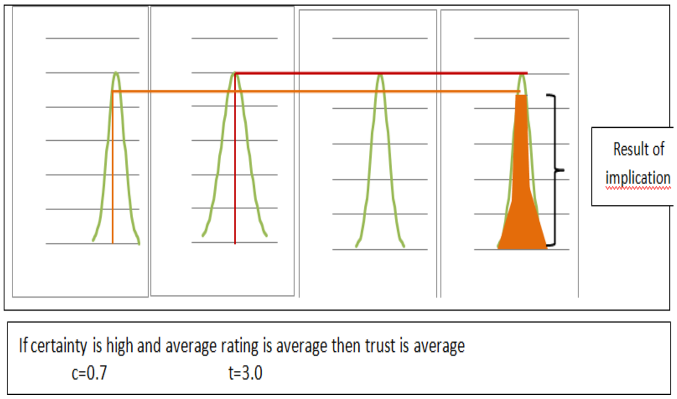

Result of implication
If certainty is high and average rating is average then trust is average
c=0.7
t=3.0

Figure 7. Implication for R14

G. *Aggregate all Outputs*

Aggregation is the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set. Aggregation only occurs once for each output variable.

It is the second last phase of defuzzification. The input of the aggregation process is the list of truncated output functions returned by the implication process for each rule. The output of the aggregation process is one fuzzy set for each output variable. In this proposed model, the fuzzy input set is certainty set and average rating set and the output fuzzy set is trust set.

H. *Defuzzification Results*

Using defuzzification equation no (9), we can get the defuzzified output. According to it, the defuzzified output is:-

$$y' = \frac{1783.81}{32.851} = 54.3 \quad (10)$$

And it continues.

I. *Mapping Surface*

In this map, we plot certainty, c and average rating, t and Trust, T. after plotting this, we get the following surface.

V. CASE STUDIES

In this section, we are going to show the impact of newly defined operators over the operators of CTM. Following this, we will show the impact of our model when it is applied is a private server or cloud. Two cases are described here, case 1 for multiple servers and case 2 for single server.

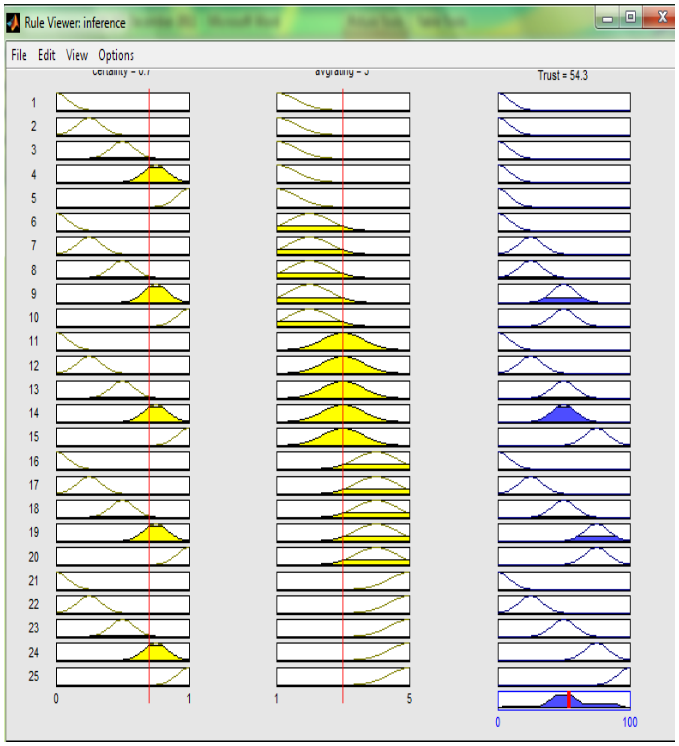


Figure 8. Sequential Process of Aggregating all outputs

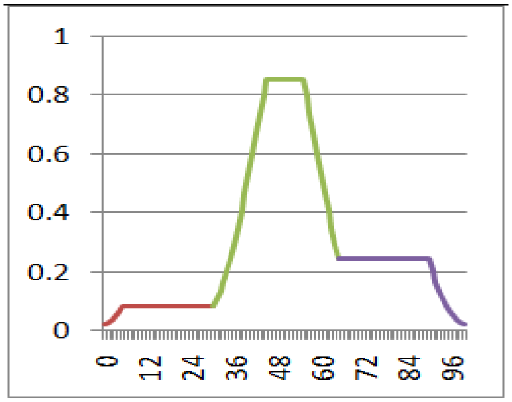


Figure 9. Defuzzified Output After Aggregation

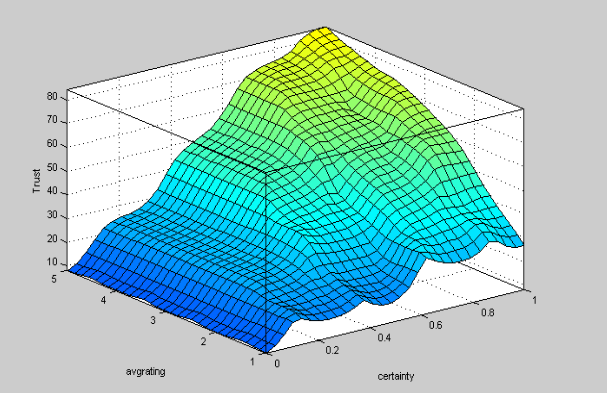


Figure 10. Defuzzified Output After Aggregation

Case 1: According to CTM's operators defined in equation 3, 4, and 5, we know that, the input for this model is r, s, f and w. Let, the input values are r=5, s=2, f=0.5 and w=1. Here, no of evidences are N=7.Then, the output values are:- average rating t = 0.714 and c=0.724. and then, E = 0.65. Now, for mapping it to our proposed model, we need to modify t. Here,

$$t' = t*\text{scale of rating} \qquad (11)$$

Usually, the scale of rating is 5. Now, the new average rating is t = 0.714*5 = 3.57. Then, the value of parameter Trust, T = ((3.57*0.724)/5)*100 = 51.69%. From fig…, we see that, it is an average situation of Trust. As the range of trust varies from person to person, one can consider it under high trust region. From the result of E and T, we can see that, it is easier to understand the condition of system or server much better that the past. As it is represented in percentage form, the user can easily understand the evaluated value of trust. Now, the value of second parameter, behavioral probability, p = 3.39% and because of T>f, the system is now in the condition of showing 3.39% higher behavior than the initial expectation. As it is in higher condition, so, the hosting partner can easily take the decision to host in this server/system. This parameter will also be useful for the developers so that they can easily understand the present condition of the system.

Now, for the system shown in figure 2 and with the help of equations given in Table No 1, we have seen the following situation: - (considering the above described values as system A)

TABLE VI. OUTPUT OF CTM AND PROPOSED MODEL

System	Let the values	Output for Certain Trust Model	Output for our proposed model
A_1***	t_{A1}=0.714, c_{A1}=0.724,	E_{A1}=0.65	T_{A1} = 51.69,M, P_{A1}= 3.39 higher
A_2***	t_{A2}= 0.459, c_{A2}= 0.806,	E_{A2} =0.467	T_{A2} = 37, M/L, P_{A2} = 26 lower
B_1***	t_{B1}=0.604, c_{B1}= 0.786,	E_{B1}= 0.582	T_{B1} = 47.47, M, P_{B1} = 5 lower
B_2***	t_{B2}=0.867, c_{B2}= 0.648,	E_{B2} = 0.74	T_{B2} = 56.18, M, P_{B2} = 12.36 higher
S_1	t_{S1}=0.829,c_{s1}=0.839, f_{s1}=0.75,	E_{s1}=0.82	T_{S1} = 69.55, H, P_{S1} = 7.26 lower
S_2	t_{S2}=0.892, c_{s2}=0.863, f_{s2}=0.75,	E_{s2}=0.87	T_{S2} = 77, H, P_{S2} = 2.67 higher
S	t_S=0.736, c_s=0.853, f_s=0.5625,	E_s=0.753	T_S = 62.78, M, P_A = 11.61 higher

**** for all cases, f = 0.5.*

Case 2: Our proposed model can also be applied for a website hosted in single server and calculating trust for a product in online transaction. Like [31], each of its portion e.g. existence, policy, fulfillment and affiliation can be represented in certainty and average rating format. And from that, we can easily calculate a website's or product trust.

Let, one person makes a target to buy a product through online. For this, that person must want to check the trustworthiness of that website from where he is going to start his business, peoples review and the quality of the product and people review of that product. One can also want to check the number of sold of that product. With our proposed model, it can easily be designed. The person, who wants to buy a product through online, can get information about it from the closer persons or someone who has an experience about the product or about the website. If it is, then that put 1 for certainty, c. That means, c=1.0.

If he is fully new to that website or product, then c=0. At the time of buying that product from website, he can give a rate to that website from different axis, such that: - served information about the product, easiness, service, privacy information, physical existence of that product, registration process, whether the website deals with updated product or not and so on.

Then, taking the average of the rating of these factors, we can get the average rating of that website. Let, average rating, t =3.75. Then, according to our model, if, c = 1 for that person, then, trust, T = 75%. It is given by that person only. Now, according to the developers of that website, initial expectation f and ranging of certainty is defined. If the developers want to take minimum 20 people's certainty, means want to take response of c=1 from at least 20 people for their system's or website's accuracy, then they can scale the certainty parameter c in following way: -

TABLE VII. Scaling of certainty for 20 people

Range of people	0<n<=4	4<n<=8	8<n<=12	12<n<=16	16<n<=20
Certainty, c	0.2	0.4	0.6	0.8	1.0

Now, for this, we can get the values for T shown in table VIII; applying the above table of values for 20 people of the organization we have explained earlier:-

TABLE VIII. FAM for representing trust (20 people)

c t	1	2	3	4	5
0.0	N	N	N	N	N
0.2	VL	VL	VL	VL	VL
0.4	VL	VL	L	L	L
0.6	VL	L	L	M	M
0.8	VL	L	M	H	H
1.0	VL	L	M	H	VH

Here, for average rating t, we have shown only scaled values. These values can also be accepted for fractional values of average rating. The rule matrix will be needed to design in the same way for those fractional average ratings.

If, the developers have a good confidence about their hosted website and after checking different mandatory requirements for that website, e.g. ensuring security level, payment method, delivery system, etc, they give the value of initial expectation, f = 0.5, then, behavioral probability, p = 50% upper. This means that, the present condition of this website is 50% upper than the initial expectation of the developers.

Let, we take 100 people/experts evidence or transaction as measurement limitation for measuring trust for experimental purpose. From that, using the basis of the CTM we have got different values for certainty c, average rating t and initial expectation f.

Applying fuzzy logic according to the table II on the experimental values we have got the values for T shown in table IX, which specify the trustiness of the system.

With the help of the FAM shown above the behavioral probability of the system can easily be calculated. With the help of T and P, one trader or user can easily take decision about the trustworthiness of the system, especially for cloud computing and the trading product. For calculating a single product's trust or rating, we have tested with 20 people's comment and experience. Secondly, we have experimented with our test system shown in (figure 2) with 100 people's experience and evidences. It seems easier for the users and developers to understand the trust and behavioral probability of the system and trust value of a single product system.

TABLE IX. FAM FOR REPRESENTING TRUST (100 PEOPLE)

c t	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
0.0	N	N	N	N	N	N	N	N	N
0.1	VL	VL	VL	VL	VL	VL	VL	VL	VL
0.2	VL	VL	VL	VL	VL	VL	VL	VL	VL
0.3	VL	VL	VL	VL	VL	L	L	L	L
0.4	VL	VL	VL	VL	L	L	L	L	L
0.5	VL	VL	VL	L	L	L	L	M	M
0.6	VL	VL	L	L	L	M	M	M	M
0.7	VL	L	L	L	M	M	M	H	H
0.8	VL	L	L	L	M	M	H	H	H
0.9	VL	L	L	M	M	H	H	VH	VH
1.0	VL	L	L	M	M	H	H	VH	VH

Our proposed representational model consists of two parameters which give us benefits according to the following points of view. These comparisons are held on the basis of paper [1], [2] and [31]:-

TABLE X. Comparison Between Models

	Original Certain Trust Model	**Ecommerce fuzzy trust model[31]**	**Our proposed model**
Fuzziness	No	Yes	Yes
Behavioral parameters	No, but one can assume	No	Yes
Certainty	Yes	No	Yes
Probabilistic Logic	Probability theory	No	Combination of Probability Theory and logic
Advantage	For Human interaction, its HTI is easy to understand	Useful for people	Useful for people and for machine.

VI. Conclusions

In this paper, we have proposed a new extension of representational model of certain trust for the evaluation of propositional logic terms, probability and fuzziness under uncertainty. It develops the representational model of the certain trust logic. Our proposed model is more expressive and useful than certain logic because it works both for machine and human beings. The parameters of the proposed model directly show how much the system can be trusted and it can be applied not only in small systems but also large systems; especially in cloud computing field. Again, it represents the present condition of the product, website and also for the system.

We have some new idea to imply in our proposed model in future. Firstly, we want to apply evolutionary algorithm with this model to optimize and design the rules. We want to apply price comparison as a parameter for a product in our model for ensuring accurate trust measuring model for a normal e-commerce website. Secondly, more development of behavioral probability parameter so that it can directly prohibit different types of security breaking questions like Sybil attack, false rating, etc is our fourth wish. At present, this parameter works indirectly with security options. Last of all, we want to establish a newer trust model with a combination of certainty, fuzzy logic, evolutionary algorithm and so on for ubiquitous computing system like cloud computing, distributed computing, etc.

Acknowledgment

The authors wish to thank Sebastian Ries, Sheikh Mahbub Habib, Max Mühlhäuser and Vijay Varadharajan for their renownable work and thinking.

References

[1] Ries, S.: Trust in Ubiquitous Computing. PhD thesis, Technische University at Darmstadt,pp: 1-192, 2009

[2] Sebastian Ries, Sheikh Mahbub Habib, Max Mühlhäuser and Vijay Varadharajan: Certain Logic: A Logic for Modeling Trust and Uncertainty, Technical report, April 6th, 2011

[3] Sebastian Ries, Sheikh Mahbub Habib, Max Mühlhäuser and Vijay Varadharajan: Certain Logic: A Logic for Modeling Trust and Uncertainty (Short Paper), June, 2011.

[4] Sebastian Ries, "Certain Trust: A Trust model for Users and Agents ", ACM, March 2007, pp. 1599-1604.

[5] Adoption, Approaches & Attitudes "The Future of Cloud Computing in the Public and Private Sectors", A Global Cloud Computing Study, JUNE 2011.

[6] Buchegger, S., Le Boudec, J.Y.A Robust Reputation System for Peer-to-Peerand Mobile Ad-hoc Networks. In: P2PEcon 2004.

[7] Junfeng TIAN, Chao LI, Xuemin HE, Rui TIAN, "A Trust Model Based on The Multinomial Subjective logic for P2P Network", International J. communications, Network and Systems, 2009, pp. 546-554

[8] ENISA: An SME perspective on cloud computing - survey. Technical report, ENISA (2009)

[9] Chow,R.,Golle,P.,Jakobsson,M.,Shi,E.,Staddon,J.,Masuoka,R.,Molina,J. : Controlling data in the cloud: outsourcing computation without outsourcing control. In: Proceedings of the 2009 ACM workshop on Cloud computing security. CCSW '09, ACM (2009) 85–90

[10] Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R.H., Konwinski, A., Lee, G., Patterson, D.A., Rabkin, A., Stoica, I., Zaharia, M.: Above the clouds: A berkeley view of cloud computing. University of California, Berkeley (Feb 2009)

[11] D.W. Manchala, "E-Commerce Trust Metrics and Models", IEEE Internet Computing, March-April 2000, pp.36-44.

[12] Mohammed Alhamad, Tharam Dillon, and Elizabeth Chang A Trust-Evaluation Metric for Cloud applications *International Journal of Machine Learning and Computing, Vol. 1, No. 4, October 2011.*

[13] Teacy, W., et al.: Travos: Trust and reputation in the context of inaccurate informationsources. Aut.Agentsand Multi-Agent Systems, vol.12, no.2, pp. 183-198, 2006.

[14] J_sang, A., McAnally, D.: Multiplication andco-multiplication of beliefs. International Journal of Approximate Reasoning vol. 38 no. 1,pp. 19-51, 2005.

[15] Jøsang, A.: A logic for uncertain probabilities. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems9(3) (2001) 279–212

[16] TJ Ross, "Fuzzy Logic with engineering applications", Wiley Online Library, 2005.

[17] Lotfi A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic", Science Direct, 13 May 1998.

[18] Vojislav Kecman, "Learning and Soft computing", Google online library, 2001.

[19] Y. Lin, W.J. Zhang, C. Wu, G. Yang, J. Dy, "A fuzzy logics clustering approach to computing human attention allocation using eyegaze movement cue", International Journal of Human-Computer Studies, Volume 67, Issue 5, May 2009, Pages 455-463

[20] Kerr, R., Cohen, R.: Smart cheaters do prosper: defeating trust and reputation systems. In: AAMAS '09: Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems, (2009) 993–1000

[21] Cristiano Castelfranchi, Rino Falcone, Giovanni Pezzulo, "Trust in information sources as a source for Trust: A fuzzy Approach ", ACM, 2003, pp- 89-96

[22] Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service Provision. Decision Support Systems 43(2) (2007) 618–644

[23] Elizabeth J.Chang, Farookh Khadeer Hussain, Tharam S. Dillon "Fuzzy Nature of Trust and Dynamic Trust Modeling in Service Oriented Environments", ACM, SWS'05, November 11, 2005, Fairfax, Virginia, USA.

[24] Florian Skopik, Daniel Schall, Schahram Dustdar, "Trustworthy Interaction Balancing in Mixed Service-oriented Systems", ACM 2010, SAC'10 March 22-26, 2010, Sierre, Switzerland

[25] L. Zadeh. "Fuzzy sets", Journal of Information and Control, 8:338—353, 1965

[26] Fuzzy Logic Fundamentals, chapter 3, pp-61-84, March, 2001

[27] Shrija Rajbhandari, Omer F. Rana and Ian Wootten, School of Computer Science, Cardiff University, Cardiff, U.K," A Fuzzy Model for Calculating Workflow Trust using Provenance Data", ACM, 2008, Baton Rouge, USA.

[28] Martin Chun-Sheng Cheng, Dynamical near optimal training for interval type-2 fuzzy neural network with genetic algorithm, thesis paper, pp-1-30, 2003

[29] .N.N. Karnik, J.M. Mendel, and Q. Liang, “Type-2 fuzzy logic system”, IEEE Trans. on Fuzzy Syst., vol. 7 no. 6, pp. 643-658, Dec 1999.

[30] Nils J. Nilson, “Probabilistic Logic*”, Artificial Intelligence 28, Elsevier Science Publishers B.V., pp – 71-87

[31] Samia, Nefti, FaridMeziane, KhairudinKasiran, A Fuzzy Trust Model for E-Commerce, 2004

[32] Ries, S., Heinemann, A.: Analyzing the robustness of CertainTrust. In: 2nd JointTrust and PST Conf. on Privacy, Trust Management and Security, pp. 51-67, 2008.

[33] Li Xiong, Ling Liu, A Reputation-Based Trust Model for Peer-to-Peer e Commerce Communities, 2003.

AUTHORS PROFILE

Kawser Wazed Nafi is now working as Lecturer in Computer Science and Engineering department of Stamford University, Bangladesh. He completed his graduation from Khulna University of Engineering and technology in Computer Science and Engineering Department. He is very ambitious and wanted to work with artificial intelligence, trust models and security in networking systems. His favorite research field is cloud computing, distributed computing, parallel computing, network security and so on.

Tonny Shekha Kar is now working as Lecturer in Computer Science and Engineering department of Stamford University, Bangladesh. She completed her graduation from Khulna University of Engineering and Technology in Computer Science and Engineering department. Her research interest is cloud computing, security and trust issues in networking systems.

MD. Amjad Hossain completed his graduation from Computer Science and Engineering department of Khulna University of Engineering and Technology and became lecturer of that department. He is now completing his PhD in Kent State University, USA. His main research interests are Cloud Computing, Quantum Computing, Image Processing, VLSI design and so on. He has many publications which are published many good quality journals.

M. M. A. Hashem received the Bachelors degree in Electrical and Electronic Engineering from Khulna University of Engineering and Technology (KUET), Khulna, Bangladesh in 1988, Masters degree in Computer Science from Asian Institute of Technology (AIT), Bangkok, Thailand in 1993 and PhD degree in Artificial Intelligence Systems from the Saga University, Japan in 1999. He is currently a Professor in the Dept. of Computer Science and Engineering, Khulna University of Engineering and Technology (KUET), Bangladesh. His research interest includes Evolutionary Computations, Intelligent Computer Networking, Wireless Networking, Soft Computing, Evolutionary Cluster Applications to Evolutionary Robots, Series: Studies in Fuzziness and Soft Computing, Vol. 147, Springer Verlag, Berlin/New York, ISBN: 3540-20901-8, (2004).

Scalable and Flexible heterogeneous multi-core system

Rashmi A Jain,
Electronics engineering department
G.H.Raisoni College of Engineering Nagpur
(M.S.) India

Dr. Dinesh V. Padole
Electronics engineering department
G.H.Raisoni College of Engineering
Nagpur (M.S.) India

***Abstract*—Multi-core system has wide utility in today's applications due to less power consumption and high performance. Many researchers are aiming at improving the performance of these systems by providing flexible multi-core architecture. Flexibility in the multi-core processors system provides high throughput for uniform parallel applications as well as high performance for more general work. This flexibility in the architecture can be achieved by scalable and changeable-size window micro architecture. It uses the concept of execution locality to provide large-window capabilities. Use of high memory-level parallelism (MLP) reduces the memory wall. Micro architecture contains a set of small and fast cache processors which execute high locality code. A network of small in-order memory engines use low locality code to improve performance by using instruction level parallelism (ILP). Dynamic heterogeneous multi-core architecture is capable of reconfiguring itself to fit application requirements. Study of different scalable and flexible architectures of heterogeneous multi-core system has been carried out and has been presented.**

Keywords-Flexible Heterogeneous Multi Core system (FMC); instruction level parallelism, thread-level parallelism; and memory-level parallelism; scalable; chip multiprocessors (CMP).

I. Introduction

A multi-core system is a single computing component. It has two or more processors which are independent of each other and each is called as a core. Each core reads and executes program instructions.

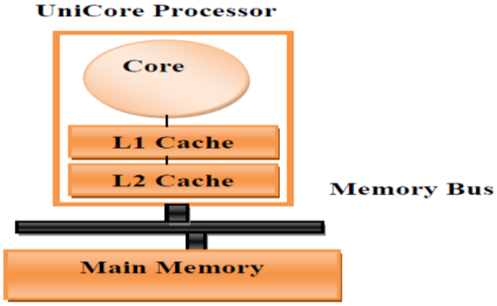


Fig. 1 Uni-core systems

The multiple cores can run multiple instructions at the same time that causes increase in overall speed of program execution. Multiple-core processors are also called as single-chip multiprocessors or more simply chip multiprocessors (CMP).

They appear similar to a traditional symmetric multi processor (SMP) but with all of its processors located on a single chip. In processors of today, there are typically 2 or 3 levels of on-chip cache

The level 1 (L1) cache is neighboring to the processor execution core and has the direct access time but the lowest capacity. Thelevel2 (L2) cache is next in the cache hierarchy and has longer access times but larger capacity. Finally, there may be a level3 (L3) cache with even longer access times but even larger capacity. On multi-core processors, the last-level cache, which is the level before requiring off-chip main memory access, is usually shared among more cores. Typically, this component is the L2 cache or the L3 cache.

Designer integrates the cores onto a single integrated circuit known as a chip multiprocessor or CMP. Processors were made with only one core. A many-core processor is also multi-core processor. In multi-core systems, the term multi-CPU refers to multiple physically separate processing-units. The many -core and multi-core are sometimes used multi-core architectures with a high number of cores tens or hundreds.

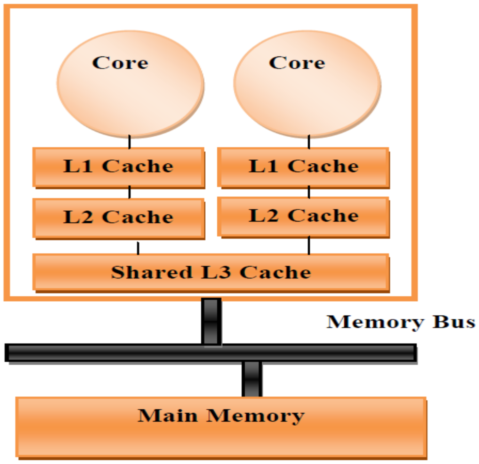


Fig. 2 Multi-core systems

A dual-core processor has twice cores a quad-core processor have four cores, it means it is made by four core a hexa-core processor have a six cores and similar like an octa-core processor have eight cores. We propose a micro architecture capable of running a single thread or many threads with high performance.

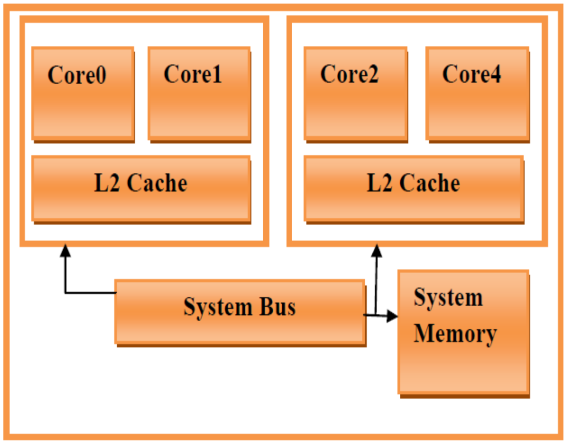


Fig.3 multi-core processor

It may couple cores in a multi-core device tightly or loosely. For example, cores may possibly or may not be share caches, and they may apply message passing or shared memory inter-core communication methods. General network topologies to interconnect cores consist of bus, ring, two-dimensional mesh, and crossbar.

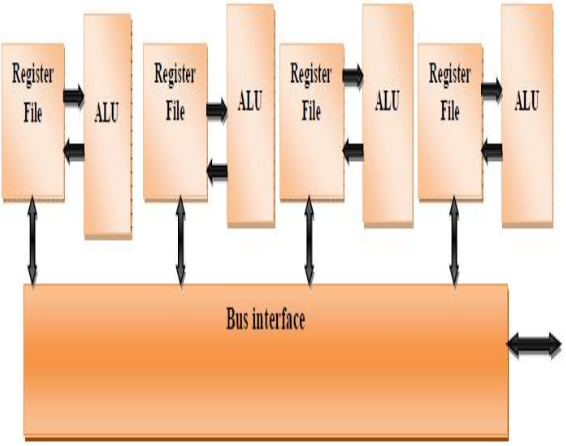


Fig. 4 Multi-core Architecture

A. *Types of multi-core system*

1. Homogeneous multi-core systems-Homogeneous multi-core systems have only identical cores. Some system use one core design repeated consistently known as homogeneous multi-core systems.
2. Heterogeneous multi-core systems-Heterogeneous multi-core systems have cores which are not identical. It means use a mixture of different cores known as heterogeneous multi-core systems.

B. *There is two kinds of heterogeneous multi-core processor*

1) Fixed heterogeneous multi-core processor-Fixed heterogeneous architecture in which partitioning remains static and it only roughly fits application requirements.

2) Flexible heterogeneous multi-core processor-There is dynamic heterogeneous multi-core architecture capable of reconfiguring itself and fit application requirement without programmer interference.

C. *Advantages of multi-core system*

1) Multitasking: -Each system has two processing cores for a maximum of twice the operating power and for better multitasking. Major advantages of multi-core system are heavy multitasking.

2) Application Support: - New applications are take advantage of this technology by using a technique known as Multithreading.

3) Power saving: - Multi-core systems have the ability to turn off one of their cores when application demand is low to save power.

Through the rest of the paper we will describe the micro architecture of our multi-core approach.

II. Flexible Micro-Architecture (FMC)

The basic of this architecture is a scalable, variable-size window micro-architecture that uses the concept of execution locality to provide large-window capabilities.

A). ILP (Instruction-level parallelism)-It calculate the many operations in a computer program can be performed simultaneously. B).TLP (Thread level parallelism) -It is a form of parallelization of computer code across multiple processors in parallel computing. c). MLP (Memory Level Parallelism) - In computer architecture the ability to have pending multiple memory operations. In a lone (one) processor, MLP may be measured a form of ILP, instruction level parallelism.

Micro-architecture uses ILP by having an effective instruction window of thousands of instructions spread across the processing elements, largely overcoming the unhelpful effects of long-latency memory operations. And it is also uses TLP for comparable workloads by allowing multiple threads to automatically allocate the processing elements. It needs to realize the greatest performance, quite than giving each thread the same kind of core of its needs. These advantages are obtained lacking changes to the ISA, compiler or operating system.

The fundamental property of the FMC is its ability to Change the instruction window size at runtime. It is able to do so by with dynamism adding or removing memory engines from the system. This property allows the processor to get used to the requirements of the application and activate only those Memory engines that are lead to improved performance.

Scalable multi-core architecture consists of a set of Cache Processors, every one with a static partition of memory engines, and a group of memory engines that can be dynamically assigned to the different threads. Figure 5 shows a general view of this micro architecture.

The architecture of FT64-3 processor contains one scalar core, one stream core, one 512KB shared secondary level cache and one DDR2 (direct device register) memory controller. Multiple processors can be connected by network interface directly. Scalar core can issue multiple instructions concurrently, and has two float pipelines, independent L1I-cache and L1Dcache, and also supports data consistency with other processors. Scalar core is responsible for running as, processing basic operations, scheduling instruction stream and data streams for the stream core, and also managing communication with off-chip. The stream core, derived fromFT64-2processor, serves as an accelerator of scalar core. Its architecture is close to that Of Imagine stream processor , which contains 16 sets of double precision FPU (grouped as 4 cluster),µc(micro controller), Stream Register File (SRF), Stream Controller(SC) etc. It is responsible for executing stream instructions received from scalar core, and moves stream data on demand. Under the control of microcontroller, 4 clusters run by means of SIMD (single instruction multiple data). After stream data is loaded into SRF, SRF transmits stream data among clusters, µc etc, and writes results back to L2Cache.

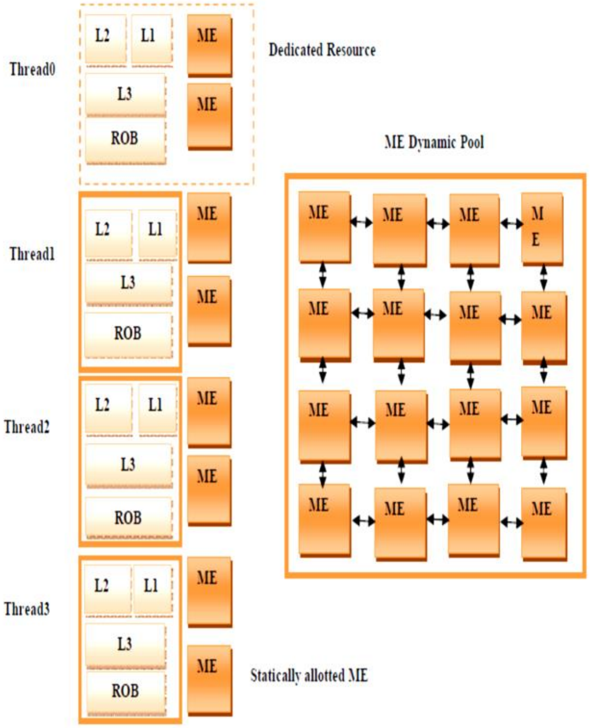


Fig.5 the micro-architecture of the flexible multi-core micro-architecture

Including a set of Cache Processors, 3 statically assigned ME (memory engine) per thread, and a dynamic pool of memory engines.

A. Memory architecture in processor core

The internal memory of scalar core is composed of common register file, 8KB Ll I-cache and 8KB Ll D-cache an it is similar to memory of traditional processor. It has execution model to uses a MLP (memory level parallelism) can tolerance latency sensitivity and bandwidth sensitivity.

III. Related Works

In addition, this design features only two execution modes that is 1.Small window or 2.Full window.

This makes it flexible chip multiprocessor. In this use the decoupled nature of this approach but overcome its limitations. Allow it to scale too many cores and many threads. The result is a processor with a variable window/issue size using a simple scalability mechanism.

Variable-size window processor .It uses multiple small cores, called memory engines. Linked through a system, to compute memory dependent instructions.

The network introduces (reduce the latency) latency, but this additional latency has little impact on instructions already waiting hundreds of cycles due to a cache miss. The memory engine network (different method for sharing the threads) can then be shared among threads to build a reconfigurable heterogeneous multi-core architecture. [1]

They proposed a new micro architecture that significantly improves performance by overcoming memory latencies while keeping complexity within reasonable limits. And they proposed a scalable micro architecture with a variable window size that can be tuned by adding or removing memory engines. They proposed a multi-threaded implementation of the micro architecture, the first heterogeneous multi-core architecture that adapts dynamically to the requirements of the threads.

Heterogeneous multi core processor can combine qualities of different architecture; it can reach peak performance as high as processors with unique architecture, though keeping as flexible as established general purpose processors at the same time. In this equivalent stream memory sub-system architecture for FT64-3 is presented. [2] It is proved that the LLC (last level cache) miss penalty is a better metric to achieve scheduling on heter-CMP system.

Hardware performance counters used to monitor the LLC miss penalty are proposed and implemented in multi-core Godson-3 RTL and simulator. An algorithm is proposed and implemented that could recognize the application behaviors accurately and schedule them to suitable cores. [3]

Additional transistors and slower clocks means multi core designs and more parallelism required .For established processor design – increasing transistor density, speeding up clock rate, and lowering voltage have been blocked by a set of physical barriers: over heat produced, also much more power consumed and also much energy leaked, useful signal reduced by noise. Multi core designs are an accepted reaction to this condition. [4]

This architecture using many copies of the same core due to this improved total computational facility on single chip.Multi-core processors have enhanced performance and area characteristics than difficult single-core processors. [5] They propose and evaluate single-ISA heterogeneous multi-core architectures as a system to reduce processor power dissipation.

They present a technique for developing dense linear algebra algorithms that seamlessly scales to thousands of cores. It can be done with our plan called DPLASMA (Distributed PLASMA) that uses a new generic distributed Direct Acyclic Graph Engine (DAGuE).[9]

This architecture using many copies of the same core .Due to this improved total computational ability on a single chip . Multi-core processors have improved performance and area characteristics than complex single-core processors. [5] Assess single-ISA heterogeneous multi-core architectures as a method to decrease processor power dissipation.

IV. SHARING OF SHARED CACHE

The common nature of on chip caches is a property that is able to be exploited for performance gains. Data and instructions that be usually accessed by every cores in a shared method be able to exist rapidly reached by all cores.

This hardware performance feature leads to our first principle of promoting distribution in the shared cache. An operating system scheduler can select processes or threads that share data or instructions and co-schedule them to every run at the similar time within the same multi core processor so that they are able to make the most of the shared cache for sharing.

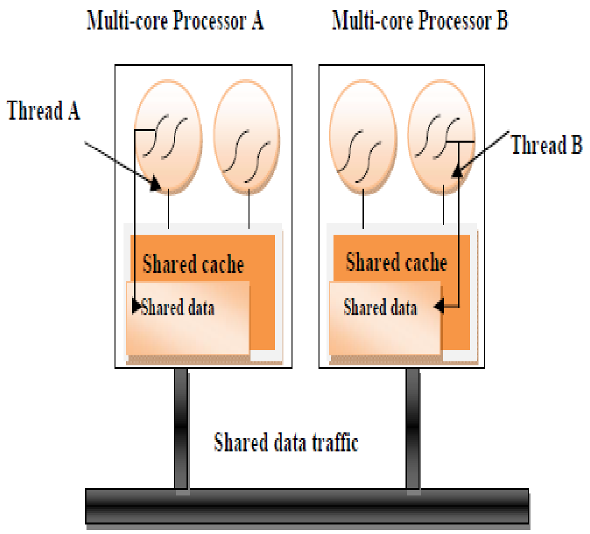


Fig. 6 Promoting sharing within a shared cache. Thread A and thread B are share data.

Thread B can be migrated to multi-core processor A so that the shared data is located within a single shared cache, resulting in more rapidly access by both threads, leading to improved performance.

A. Multiprocessor Parallelism

To enhance the number of instructions completed for every processor clock cycle, parallel instructions can be extracted from a sequential instruction stream as long as data dependencies between instructions.

If instructions B and C depend only ahead the result of instruction A, instruction D depends just upon the result of instruction B, and instruction E depends only upon the result of instruction C, then the most obtainable instruction-level parallelism for this instruction stream is 2.

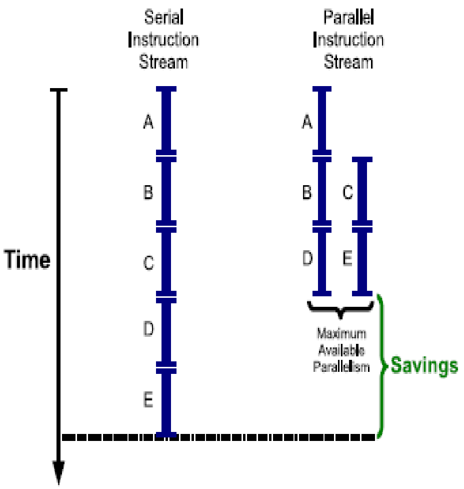

Serial Instruction Stream
Parallel Instruction Stream
Time
A
B
C
D
E
A
B
C
D
E
Maximum Available Parallelism
Savings

Fig. 7 Multiprocessor Parallelism

B. Moore's Law is Impotent

From an arithmetical point of view, performance inside a processor such as the Intel x86 family, can be calculated by the number of instructions completed per second (IPS: instructions per second), and is a result of the following.

$$\text{IPS}\left(\frac{\text{instruction}}{\text{second}}\right) = \text{IPS}\left(\frac{\text{instruction}}{\text{cycle}}\right)\ \text{clock\_frequency}\left(\frac{\text{clock}}{\text{second}}\right)$$

Given N threads .To enhances the value of IPS, and thus performance, clock frequency can be increased, or IPC can be increased. Unfortunately, clock frequency now appears capped and IPC has imperfect potential due to difficulty in extracting instruction-level-parallelism (ILP) from a single serial instruction stream. Moore's Law no longer correlates to superior clock frequency values. Moore's Law does offer extra transistors that can help enhance IPC.

C. Thread-Level Parallelism

TLP describes the situation where there are many, independent threads of execution, which can be run at the same time inside a single processor. These multiple threads can come from either inside a single application or across multiple applications. This situation is like to the traditional multiprocessor parallelism where many threads are executed on many processors. Thread-level parallelism enables the IPC term in Equation to become the sum of the IPC of each thread, as exposed in Equation for N threads.-

$$\text{IPS} = \left(\sum_{n=1}^{N} \text{IPC n}\right) \times \text{clock\_frequency}$$

D. Memory Wall

Another matter limiting performance is the huge and growing disparity between processor speeds and main memory speeds in terms of both latency and bandwidth. This disparity is usually referred to as the memory wall. It will be reducing.

E. Exploiting Multiple Processors

There are many processors, caches, main memory banks, interconnects, and I/O devices .The operating system is responsible for the smart management of these common hardware resources. These hardware performance issues are usually addressed by the operating system by scheduling and memory management.

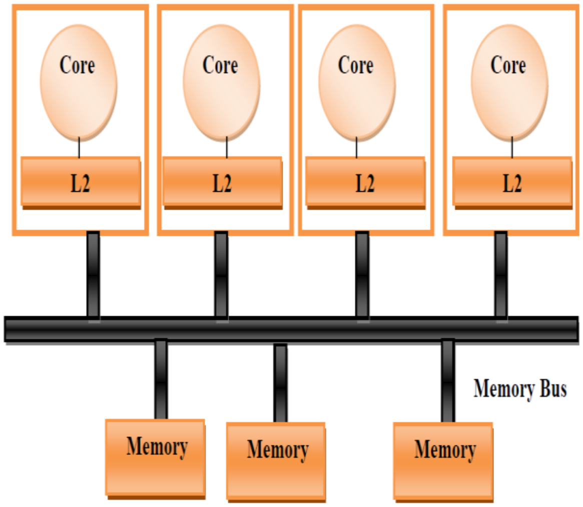


Fig. 8 SMP multiprocessor consists of several uni-core processors connected by a single shared memory bus.

F. Exploiting Simultaneous Multithreading

In simultaneous multithreading (SMT) processors, several micro-architectural hardware resources are shared among many threads of execution, leading to potential interference between threads. In adding to the processor pipeline resources, on-chip caches are too shared, which have the L1 instruction, L1 data, andL2 caches. The operating systems have to consider how to manage these shared hardware resources, by memory management and scheduling, in order to maximize application performance.

G. Exploiting Multiple Cores

On multi-core processors, the major hardware property that must be considered by the operating system is that there be able to be on-chip caches shared by many cores. This includes the L2 cache or, if it is presents the L3 cache. In difference, the L1 caches are private to each core, dissimilar in SMT processors. Another hardware property that should be considered by the operating system is that communication among cores is quicker than on traditional multi-chipped multiprocessors because every core is located on the identical chip, sharing the similar on-chip L2 cache

V. The Decoupled KILO-Instruction Processor (D-KIP)

In the D-KIP, two cores used to implement an application. The first core, the Cache Processor (CP), is small and fast, and executes all code depending only on cache accesses (high locality code). The CP runs forward as fast as possible, launching all loads with known addresses. Code to depend on memory accesses (i.e., low locality code) is processed through a secondary core, the Memory Processor (MP), which is proposed as a small in-order processor.

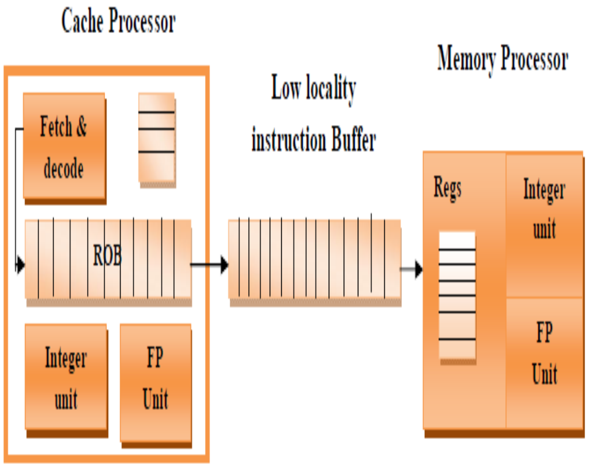


Fig. 9 Micro architecture of the D-KIP Processor

It executes low-locality code fetched from an in-order Low Locality Instruction Buffer (LLIB) that has earlier been filled in program order through the CP. Processor recovery is ensured by using checkpoints that are formed dynamically at the reorder buffer (ROB) output of the CP. Figure 9 shows a simplified overview of the D-KIP processor.

VI. Resizable Window with a Set of Memory Engines Memory Management

To partition the buffer into many in order smaller Buffers and provide each one with its own set of functional units. The buffers are then allocated round-robin to the cache

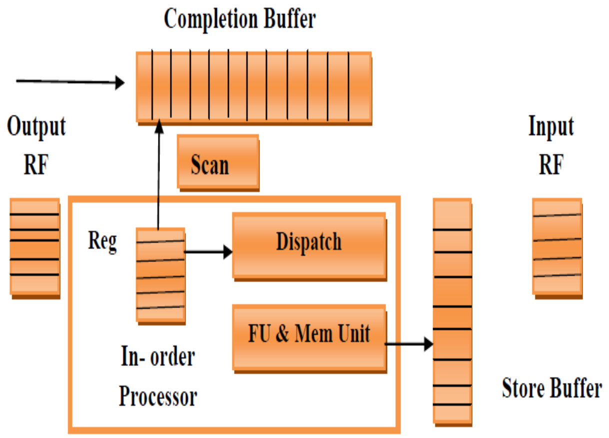


Fig.10 Architecture of a single Memory Engine

Processor as they are desired. In this scheme, instead of having a single memory engine we have a memory Processor consisting of a set of Memory Engines. Figure no.10 shows the Architecture of a single Memory Engine. Each of these memory engines is a copy of the multi-scan engine behavior a subset of the compressed instruction window.

Registers are passed between the engines with the input and output register files. After each scan the engine checks the output register file for recently generated registers .These registers are at that time sent over a network to the input register file of the next logical memory engine. Memory management is a critical part in high performance architectures.

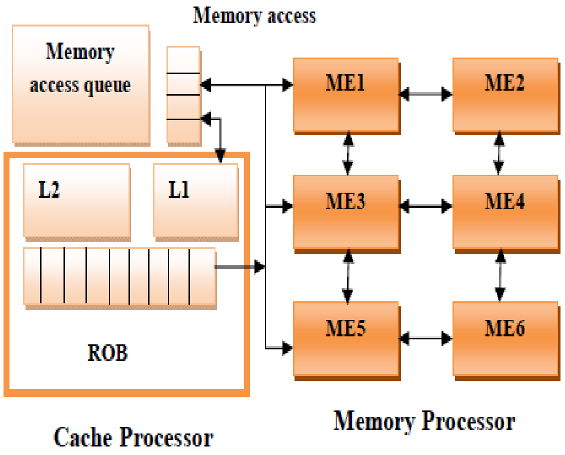


Fig.11 Processor with ME Network

VII. Conclusion

We presented a flexible multi-core (FMC) micro architecture able of with high performance. It high throughput used for identical parallel applications as well as high performance. Each core is extremely simple, thus our approach scales to a large number of cores, allowing for a capable and easy design. We consider this according to every method it gives the right path to provide best performance for workloads consisting of a large variety of applications. FMC performs this transparently to the programmer.

References

[1] Miguel Pericas, Adrian Cristal, Francisco J. Cazorla, Ruben Gonzalez, Daniel A. Jimene and Mateo Valero "A Flexible heterogeneous Multi-Core Architectura".20 Parallel Architecture and Compiler Techniques 2007.

[2] Rakesh Kumar et. Al, "Single-ISA Heterogeneous Multi-Core Architectures: The Potential for Processor Power Reduction", In Proceedings of the 36th International Symposium on Micro architecture, December 2003.

[3] Rangyu Deng et. "An Efficient Stream Memory Architecture for Heterogeneous Multi core Processor 2009.

[4] Shouqing Hao et. Al "Processes Scheduling on Heterogeneous Multi-core Architecture with Hardware Support" 2011.

[5] George Basilica et. Al "Flexible Development of Dense Linear Algebra Algorithms on Massively Parallel Architectures with DPLASMA"2011.

[6] H. Akkary, R. Raj war, and S. T. Srinivasan. Checkpoint processing and recovery: Towards scalable large instruction window processors. 2003.

[7] S. Balakrishnan, R. Raj war, M. Upton, and K. Lai. The impact of performance asymmetry in emerging multi core architectures. In Proc. of the Intl. Symp. On Computer Architecture, pages 506–517, June 2005.

[8] R. D. Barnes, S. Ryoo, and W. Mei W. Hwu. Flea-Flicker Multipass Pipelining: An Alternative to the High-Power Out-of-Order Offense. In Proc. of the 38th. Annual Intl. Symp. On Micro architecture, December 2005.

[9] Miguel Prices`, Adrian Cristal, Ruben Gonzalez, Daniela. Jimenez and Mateo Valero "A Decoupled KILO–Instruction Processor "in 2005.

[10] Miguel Prices et all" Chained In-Order/Out-of-Order Double Core Architecture" in 2006

[11] Francisco Javier Cazorla Almeida "Quality of Service for Simultaneous Multithreading Processors" in 2005.

[12] O. Mutlu, J. Stark, C. Wilkerson, and Y. N. Patt. Run ahead Execution: An alternative to very large instruction windows for out-of-order processors. In Proc .of the 9th Intl. Symp .on High Performance Computer Architecture, pages 129–140, 2003.

[13] M. Pericas, A. Cristal, R. Gonzalez, D. A. Jimenez, and M. Valero. A decoupled kilo-instruction processor. In Proc. of the 12th Intl. Symp. on High Performance Computer Architecture, February 2006.

[14] H. Zhou. Dual-core execution: Building a highly scalable single-thread instruction window. In Proc. of the 14th Intl. Conf .on Parallel Architectures and Compilation Techniques, September 2005.

[15] H. Akkary, R. Rajwar, and S. T. Srinivasan. Checkpoint processing and recovery: Towards scalable large instruction window processors. 2003.

[16] R.Kumar, V.Zyuban, and D.M.Tullsen. Interconnection in multi-core architectures: Understanding mechanisms, overheads, an scaling. In Proc of the 32ndIntl.Symp. On Computer Architecture, June 2005.

[17] John Nickolls, Ian Buck, Michael Garland, and Kevin Skadron. Scalable parallel programming with CUDA. ACM Queue, 6(2):40{53, 2008.

[18] M. D. Lindeman, J. D. Collins, H. Wang, and T. H. Meng. Merge: a programming model for heterogeneous multi-core systems. In ASPLOS XIII, 2008.

Genetic Algorithm Based Approach for Obtaining Alignment of Multiple Sequences

Ruchi Gupta[1]
Ph.D Research Scholar MCA Deptt
AKGEC Ghaziabad

Dr. Pankaj Agarwal[2]
Professor & Head, Department of
Computer Science & Engineering
IMS Engineering College Ghaziabad

Dr. A. K. Soni[3]
lProfessor & Head, Department of
Computer Science & Engineering
Sharda University, Greater Noida

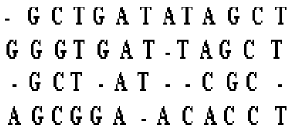
***Abstract*—This paper presents genetic algorithm based solution for determing alignment of multiple molecular sequences. Two datasets from DNA families Canis_familiaris and galaxy dataset have been considered for experimental work & analysis. Genetic operators like cross over rate, mutation rate can be defined by the user. Experiments & observations were recorded w.r.t variable parameters like fixed population size vs variable number of generations & vice versa, variable crossover & mutation rates. Comparative evaluation in terms of measure of fitness accuracy is also carried out w.r.t existing MSA tools like Maft, Kalign. Experimental results show that the proposed solution does offer better fitness accuracy rates.**

Keywords-DNA Sequences; alignment; Genetic Algorithm; Crossover; Mutation; Selection; Multiple Sequence Alignment etc.

I. Introduction

Simultaneous alignment of several sequences is among the most important problems in computational molecular biology. Multiple sequence alignment (MSA) can be seen as a generalization of Pairwise Sequence Alignment where instead of aligning two sequences, n sequences are aligned simultaneously, where n is > 2. Multiple sequence alignment can discover biologically significant sequence patterns that may be widely dispersed or hidden in the molecular sequence databases. MSA gives insight into the basis for sequence of similarities between homologous sequences. [1]

An example of an alignment of four hypothetical DNA sequences is shown in Fig. 1.



```
- G C T G A T A T A G C T
G G G T G A T - T A G C T
- G C T - A T - - C G C -
A G C G G A - A C A C C T
```

Figure1: An Example of an Alignment

The basic idea is that the sequences are aligned on top of each other, so that a coordinate system is set up, where each row is the sequence for one protein, and each column is the 'same' position in each sequence. Each column corresponds to a specific residue in the 'prototypical' protein.

Multiple Sequence Alignment (MSA) is considered to be an important tool for computational biologists. It finds its application in phylogenetic analysis, identification of conserved motifs and domains and structure prediction [3]. MSA is a computationally difficult problem, also known to be a NP-hard problem [2]. Considering both the importance and complexity of solving the MSA problem, many different heuristic methods have been proposed by the researchers to provide approximate solutions to this problem.

Genetic Algorithms (GAs) as a computational means to solve the MSA problem has shown lot of potential. It can search through the solution space effectively and generate good alignment results. The main advantage of genetic algorithms over other optimization methods is that there is no need to provide a particular algorithm to solve a given problem. It only needs a fitness function to evaluate the quality of different solutions. Also since it is an implicitly parallel technique, it can be implemented very effectively on powerful parallel computers to solve exceptionally demanding large-scale problems.

The method works by breaking a series of possible MSAs into fragments and repeatedly rearranging those fragments with the introduction of gaps at varying positions. This paper also explores the possibility of applying GA based solution for MSA problem. One such proposed & developed solution is also presented.

II. Related Study

Genetic algorithm is one of the useful tools determining alignment of multiple sequences. Iterative methods may be implemented through evolutionary approach that use computational heuristics based on natural biological phenomena such as selection, crossover and mutation to evolve a population of candidate solutions based on an objective function because they work similarly to progressive methods but repeatedly realign the initial sequences as well as adding new sequences to the growing MSA [3].

There are some proposed iterative methods to improve the problem of MSA. For example, evolutionary approach SAGA [5] based on genetic algorithm have been success fully applied to the MSA problem. It is used to optimize two different objective functions and shows that they can search large solution space efficiently. But due to repeated use of fitness function it may increase its time complexity.

Zhang C et al., [7] proposed an algorithm based on genetic algorithm and dynamic programming. It was used with two different distance matrices and characterized by great complexity in processing time. It has some limitations for performing crossover and mutation operations.

One of the most appropriate GA approaches to solve the MSA problem was presented by Nguyen et. al [8], however there are still some limitations w.r.t scoring scheme.

Another useful algorithm for multiple DNA sequence alignment using genetic algorithms and divide-and-conquer techniques [9] was proposed in which optimal cut points of multiple DNA sequences were selected. According to the author experimental results showed quite significant results. Approach involves cutting of the sequences for decreasing the space complexity for sequence alignment. However alignment was possible only for multiple deoxyribonucleic acid sequences, not for protein and other nucleic acid sequences.

Other new genetic algorithms [10] were used for solving the MSA in which various dataset were tested and the experimental results were compared with other methods. But after comparison it was observed that this approach could obtain good performance in the data sets with high similarity and long sequences.

After that effective GARS approach [11] based on Genetic Algorithm with Reverse Selection was proposed. But it suffers from premature convergence in which solution reaches locally at an optimal stage. Furthermore a new approach AlineaGA [12] was proposed which used a

Genetic Algorithm with local search optimization embedded on its mutation operators for performing multiple sequence alignment. But its mutation probability leads to better solutions in fewer generations and that the mutation operators had a dramatic effect in this particular domain. Recently a new Cyclic Genetic Approach (CGA) [13] developed with the complete knowledge of the problem and its parameters. In CGA, the values of various parameters are decided based on the problem and fitness value obtained in each generation. But the column score value varies for each execution may not give relatively better alignment.

In this paper, we proposed an evolutionary approach using genetic algorithms to obtain alignments of multiple sequences. Experimental results show that the proposed solution does offer better fitness accuracy rates w.r.t some existing tools.

Methodology

The remainder of this section is organized as follows. . In section 3 we present genetic algorithm based approach (GAMS) for solving the problem of aligning multiple sequences. Section 4 shows the experimental results of various dataset which are used to test the performance of our method. Then section 5 is finally used for discussion and conclusion.

III. GENETIC ALGORITHM BASED APPROACH

In this section we present our algorithm for solving the MSA problem. Genetic algorithms based approach (GAMS) are applied with new selection and crossover scheme which helps us to generate best population on local schema so that better alignment could be discovered. This process flow is depicted in figure 2.

```
GA_MS ()
// initialize a usually random population of individuals
        initpopulation P (t);
// evaluate fitness of all initial individuals in population
        Evaluate P (t);
// test for termination criterion (fitness core)
 While (not find best solution)
{
FOR i = 1 TO n DO
{
 // Select two chromosomes X and Y with highest fitness value from current evaluation
        P' := select parents P (t);
// recombine the "genes" of selected parents
        Recombine P' (t);
// perturb the mated population stochastically
            Mutate P' (t);
 // evaluate its new fitness
            Evaluate P' (t);
 // select the survivors from actual fitness
            P := survive P,P' (t);
        Od
       }
    end GA.
}
```

Figure 2: GA Process flow

A. *Chromosome Representation*

The chromosome should in some way that contains information about solution which it represents. The most used way of encoding is a binary string. The chromosome then could look like this: Each chromosome has one binary string. Each bit in this string can represent some characteristic of the solution or the whole string can represent a number. Of course, there are many other ways of encoding. This depends mainly on the solved problem. For example, one can encode directly integer or real numbers; sometimes it is useful to encode some permutations and so on. Each sequence has its own length. The number of gaps in the sequence is to be inserted in each sequence. It is calculated in a way that the length of all sequence remains the same. Therefore we have to generate the maximum length of sequence by multiplying the maximum length of particular element of sequences with rsp1.2. Let’s say we have a set of sequence S = {S1, S2, S3 ….Sn}. So the maximum length of the column has to be found out by multiplying the sequence with rsp by maximum length column. The value of scaling factor rsp defines that the alignment to be 20% longer then the sequence which is based on the observations that solution to common MSA problem really contains more than 20% gaps. The flow of chromosome representation is shown in figure 3:-

```
Create_Random_Matrix(n,L)
{
    //generate the initial population G
    //let n be size of Population
    FOR  G = 1 TO n DO
    {
        Select Length of Input Sequence=L
        For  i=1 to L DO
        {
            //Create Position of Random Spaces
            //Generate minimum value is 1
            //Generate Position of Space
            // define the value of scaling factor=1.2
             Position = (RandomSpaces(min, L + 1));
        }
    }
}
```

Figure 3: chromosome representation

B. Evaluation of Fitness Function

To evaluate their fitness, the chromosomes must be converted to the alignment form to be applied sum-of-pairs function [3]. We scored each column by looking at matches, mismatches, and gaps in the two sequences. We assume that a match = 1, a mismatch = 0, and a gap = -1. The fitness or scoring function of each individual is calculated by the formula:-

$$Fitness\_Score = \sum_{i=1}^{p-1} \sum_{j=j+1}^{p} ScoringMatrix(A_i, A_j)$$

The fitness Score for each alignment is calculated by summing the individual score for each column in the matrix. Scoring matrix is needed to determine the cost of aligning a residue with another. Also, a gap penalty value must be settled for determining the cost of aligning an amino acid with a gap. This penalty is only employed when aligning a residue with a gap. The fitness value calculation is to be represented by figure 4:-

```
FitnessValue(G, Max_length)
{
    //generate G is the Gap penalty
    //Calculate Sequence Count
    //Calculate max_length of Sequences
    for(a=0 to (max_length+a))
    {
        //check position of sequence in matrix not null or  ' - ' fitnessvalue+=0
        //check position of sequence in matrix  null or '-‘'
        fitnessvalue += G
    }
}
```

Figure 4: The flow for finding best fitness score

C. Selection Procedure

After calculating the fitness score of all the population applying larger tournament method where n individuals are randomly chosen, the fitter of the two is selecting with the highest and second highest fitness value .In this case the fitter the individual is chosen by the following procedure:-

- Apply larger tournament strategy for the current population based on their fitness function

- Select two best chromosomes randomly based on their column score and select two individual with their highest fitness value.

D. Crossover

In the single point crossover process, Crossover selects sequence from parent chromosomes and creates a new offspring. The simplest way how to do this is to choose randomly some crossover point and everything before this point copy from a first parent and then everything after a crossover point copy from the second parent .we select crossover point at the rate of 0.5 and count the entire gap in each population then multiply it with crossover rate and take ceiling of crossover rate. The crossover point is selected by the formula:-

$$Crossover \ldots po\text{int} = total \ldots no \ldots of \cdots gaps \times 0.5$$

After selecting point, copy the chromosome of first parent exact at the crossover point value then copy all chromosome of second parent and vice versa so [13]. There are two offspring has to be generated after applying the crossover function. Calculate the fitness score of current population and select the best individual for performing mutation operation. The flow of one point crossover is shown in figure 5.

```
CrossOver (b,p)
{
    //let cr be the cross rate
    //let crosspoint be the cp
    //cp=sequencecount+cr
    //pick an array b[] in the range crossover from random
    //Declare p as point =b[]
    for i=1 to  Do cp
    {
```

```
if b[i]=p[p+cp]Go to first step
else
b[p+cp]=p[p+cp-i]
}
//let position of matrix pos
pos=b[i]
if b[i]>max_length
max_length +=max_length
pos=pos-(max_length*(sequencecount-1))
 }
```

Figure 5: flow of one point crossover

E. Mutation

After a crossover is performed, mutation takes place. This is to prevent falling all solutions in population into a local optimum of solved problem. The system randomly chooses a gene of a chromosome form the mating pool randomly and applying binary mutation. Mutation changes randomly the new offspring. For binary encoding we can switch a few randomly chosen bits from 1 to 0 or from 0 to 1 [9]. where all the gaps are represented by 0's and all the base symbols are represented by 1's and mutation takes place separately in each sequence up to the mutation point rate of 0.2 [9] is initialized and corresponding mutation point is selected. The mutation point is to be selected by the formula:-

$$Mutation\ldots po\text{int} = 0.2 \times total\ldots length\ldots of\ldots bit \cdots string$$

First the mutation operator converts the total sequence in to bit string then calculate the mutation point after calculating the mutation point every picks a random amino acid from a randomly chosen row (sequence) in the alignment and checks whether one of its neighbors has a gap. If this is the case, the algorithms swap the symbols. The flow of one point crossover is shown in figure 6.

```
Mutation (mp,i)
{
//Declare mutation point mp
//declare sequence row count count
mp=count* mutation rate
//declare string symbol
 for i=1 to mp do
{
symbol=removeSymbols.SubString(i,1)
 if(matrix row='-')
matrix row =symbol
else
 matrix row = '-'
}
}
```

Figure 6: flow of space mutation

IV. Implementation and Results

The algorithm is implemented using Microsoft visual studio and the machine for this research is a personnel computer with Intel Pentium III processor .The main memory is 4 gigabyte and Microsoft XP was used as a platform for the implementation. The DNA query input sequence is to be taken from cans family. Query input format of the DNA sequence is listed in figure 7:-

>SPAC1002.14|1824570|1826248|itt1|I

GTAAATTCATACCGGAAATTTTACCAAATGGCGAT
TTCTTAATTGCTGAGGTGGCCAGCAGAAATCGTCTTT
TCATTATTCTGGAATCAAAACACATTCTTTGAATTGTT
CACTTTTCTGTTGCCTTGAAATCTTGGTCTTCTTAGTT
GACTGTTTCATCAAGGTTGCTCCAAATTCTTTGTGATT
TATTGGTAAACTCGGGCATTTTATTGAGT

>SPAC10F6.10|1225464|1227365|SPAC10F6.10|I

TTTTTATATACCAGTTTTATTTACAACAAAAAGTTT
TTACTACCACCTACAAAATACAAAAACTTGGATTTGT
ATCCAGTTCTTTGTCAAATTTTTAAATAAATTATTCTT
TTATTGATTTATTTAAAGTTTAAG

Figure 7: Query Input Formats

During the course of experiments, we have tried various chromosome lengths in order to understand how they have an effect on the performance of the GA.

Datasets

We have used two datasets which are DNA sequences from two DNA families, Canis_familiaris dataset (psm3, SPAC105.03c, taf11, SPAC1142.01) and galaxy dataset (AY395516.1,AY390420.1,AY390421.1,AY390422).These datasets are used as input to our multiple sequence alignment. The parameters setting for experiment are summarized in table 1.

Parameter	Content
Population size	5,10,15,20
Generation	50,100,150,200
Selection Strategy	Random Selection
Crossover operator	One Point
Crossover Rate(Rc)	0.8,0.6,0.3
Mutation operator	Space Mutation
Mutation Rate(Rm)	0.5,0.3,0.1
Scoring Matrix	Scoring 1
Gap penalty	-1
Accumulate size	20%

Table 1: GA Parameters

In order to examine our algorithm validity, we test number of series with DNA sequence. Firstly the algorithm is executed 100 run on GA while number of generation become fix and size of population is to be continued changed. It is observed that running time is increased accordingly and indicate by a notable rise in fitness score about 10% after increasing each size of population. Again while size of population become fixed and number of generation continued to changed then it has to be notified that running time is increased accordingly and indicate by a notable rise in fitness score about 10% . Table 2, 3 and figure 8, 9 lists the results after varying the number of generation and size population.

Size of Pop	No of Gen	Fitness Score	Running Time
5	100	-831	2:06
10	100	-931	4:78
15	100	-940	8:20
20	100	-990	12:08

Table 2: Operators assemble on 5, 10, 15 and 20 size of population with fix number of generation with and calculated fitness score, running time

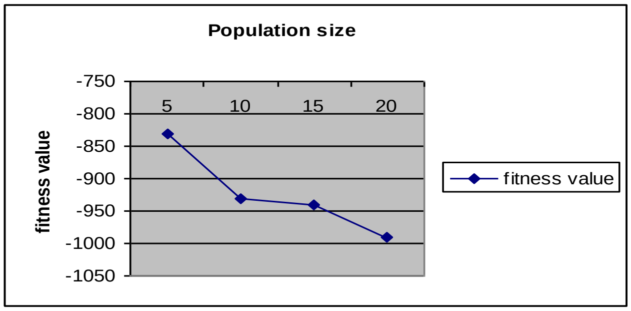


Figure8: Fitness curve GA with Verifying size of population

Size of Pop	No of Gen	Fitness Score	Running Time
5	50	-831	1:03
5	100	-889	2:05
5	150	-909	3:07
5	200	-1002	4:06

Table 3: Operators assemble on 50,100,150 and 200 numbers of generations with fix size of population and calculated fitness score, running time

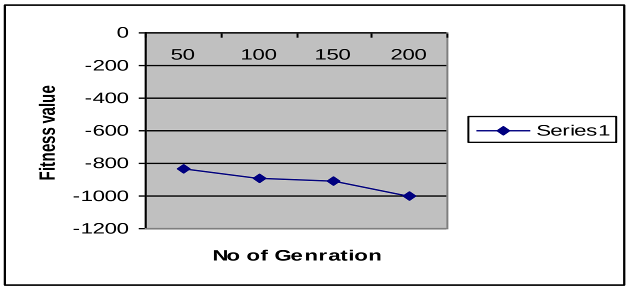


Figure9: Fitness curve GA with Verifying number of generation

In the proposed solution, we have also used two specific crossover and mutation operators. In order to determine the best crossover and mutation probabilities; we have carried out three different experiments, using ten randomly selected canis_familiaris dataset that were obtained from [15]. In our experiments for each of ten datasets, the algorithm is executed 200 run on GA and the statistical outcomes of the optimal fitness in each run is calculated as the results. We measure the best fitness score and running time for each generation. Our algorithm has to be run with the 30% crossover & 10%-mutation option, 60% crossover & 30% mutation option and 80% crossover & 50% mutation option .it is observed that our algorithm obtained the best solutions for 80% crossover & 50% mutation option .The solutions obtained by the 60% crossover and 30% mutation for the same datasets are close to the best scores, however the option 30% crossover & 10% mutation has not achieved any good quality solutions. Therefore, we can conclude that GA has achieved overall better performance for these test datasets when the rate of crossover are selected as 80% and mutation are selected as 50%. As for results for these datasets are to be presented in table 4 and corresponding plots are to be presented in graph 10.

Cros rate=30% Mut rate=10%	Cros rate=60% Mut rate=30%	Cros rate=80% Mut rate=50%	Pop Size	Running Time
-950	-940	-920	5	1:03
-935	-923	-900	10	2:05
-900	-868	-860	15	3:07

Table 4: fitness scores with selected Crossover and mutation rate options

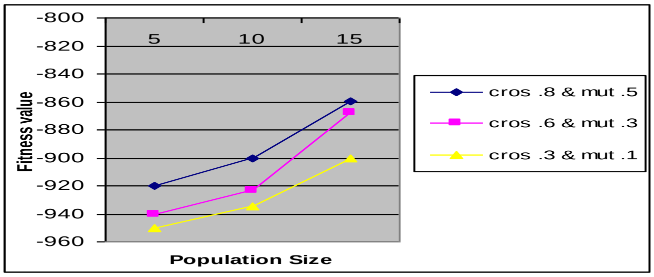


Figure10: Experimental results on GA with selected Crossover and mutation rate options

The last set of experiments compares our algorithm (GAMS) with two different tools such as Maft (high speed multiple sequence alignment program) and Kalign (fast and accurate multiple sequence alignment algorithm). The maximum 200 generation run on GA and the statistical outcomes of the optimal fitness in each run is calculated as the results. The sequence id and specification of each dataset is given in table 5. We measure that our GA obtained the best fitness score and running time for each generation as compare to other. The GA typically found a good alignment within 200 generations. Table 6 and graph 10 lists the results after comparisng our algorithm with Maft and kalign.

Seq ID	Sequence Specification	No of Sequence
A1	>SPAC1142.01 >SPAC1002.04c >SPAC105.03c >SPAC10F6.09c	4
A2	>SPAC1006.08 >SPAC1006.05c >SPAC1002.20 >SPAC10F6.08c	4
A3	>SPAC1002.02 >SPAC11E3.06 >SPAC1002.08c >SPAC10F6.03c	4
A4	>SPAC1002.13c >SPAC11D3.01c| >SPAC10F6.14c >SPAC1071.07c	4

Table 5: Canis_familiaris dataset

Seq ID	GAMS	Maft	Kalign
A1	-4679	-6446	-6795
A2	-2367	-3469	-3868
A3	-4317	-5814	-6557
A4	-3408	-5450	-6411

Table 6: Overall Performance of all methods of Sequence ID datasets

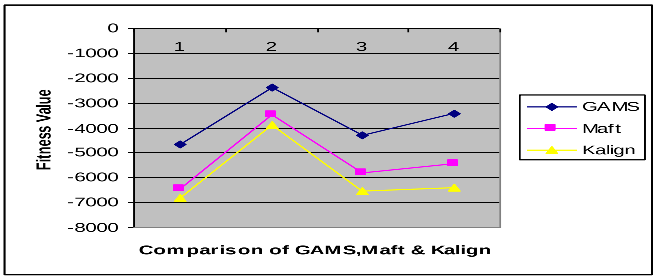


Figure10: Overall Performance of all methods of Sequence ID datasets

V. Conclusion

Multiple sequence alignmen is an extension of pairwise alignment to incorporate more than two sequences at a time. Our multiple alignment methods try to align all of the sequences in a given query set. Efficient fitness value function, crossover and mutation strategies are the outcome of work. Eventually it is trying that our methods will be significantly contributed in prior efficient solution to multiple sequence alignment problems.

Refernces

[1] C. Gondro and B.P. Kinghorn, "A simple genetic algorithm for multiple Sequence alignment", Genetics and Molecular Research 6 (4): 964-982 (2007)

[2] Kosmas Karadimitriou and Donald H. Kraft, "Genetic Algorithm and the Multiple Sequence Alignment in Biology ", Proceedings of the Second Annual Molecular Biology and Biotechnology Conference, February 1996, Baton Rouge, LA.

[3] C. Notredame, « Recent progresses in MSA a survey. , pharmacogenomic, volume 3, pages 1–14, 2002.

[4] Fernando José Mateus da Silva, Juan Manuel Sánchez Pérez, Juan Antonio Gómez , "Optimizing Multiple Sequence Alignment by Improving Mutation Operators of a Genetic Algorithm", 978-0-7695-3872-3/09 © 2009 IEEE

[5] C. Notredame and D.G. Higgins. "SAGA: sequence alignment by genetic algorithm", Nucleic Acids Research, volume 24(8): 1515–1524, 1996.

[6] L. Davis, "Handbook of Genetic Algorithms." Van Nostrand Reinhold, New York, 1991.

[7] Zhang C, Wong, "AKC: Toward efficient multiple molecular sequence alignment: a system of genetic algorithm and dynamic programming", IEEE Transactions on Systems, Man and Cybernetics, Part B 1997, 27:918 -932.

[8] Nguyen HD, Yamamori K, Yoshihara I, Yasunaga M, "Improved GA-based method for multiple protein sequence alignment", The 2003 Congress on Evolutionary Computation (CEC '03) 2003, 3:1826 - 1832.

[9] Shyi-Ming Chen, Chung-Hui Lin, and Shi-Jay Chen, "Multiple DNA Sequence Alignment Based on Genetic Algorithms and Divide-and-Conquer Techniques", International Journal of Applied Science and Engineering (2005). 3, 2: 89-100

[10] Jorng-Tzong Horng, Li-Cheng Wu, Ching-Mei Lin, Bing-He Yang, "A Genetic Algorithm For Multiple Sequence Alignment", Soft Computing-A Fusion of Foundations, Methodologies and Applications, Vol. 9, Issue 6, pp 407 – 420. (2005)

[11] Yang Chen, Jinglu Hu, Member, IEEE, Kotaro Hirasawa, Member, IEEE, Songnian Yu. "Multiple Sequence Alignment Based on Genetic Algorithms with Reserve Selection" , ICNSC, pp 1511-1516 (2008).

[12] Fernando José Mateus Silva, Member, IEEE, Juan Manuel Sánchez-Pérez, Juan Antonio Gómez-Pulido and Miguel A. Vega-Rodríguez , "An Evolutionary Approach for Performing Multiple Sequence Alignment", 978-1-4244-8126-2/10/$26.00 ©2010 IEEE

[13] Amouda Nizam,Jeyakodi Ravi1, and Kuppuswami Subburaya2, "Cyclic Genetic Algorithm for Multiple Sequence Alignment", International Journal of Research and Reviews in Electrical and Computer Engineering (IJRRECE) Vol. 1, No. 2, June 2011

[14] Guang-Zheng Zhang De-Shuang Huang,"Aligning Multiple Protein Sequence by AnImproved Genetic Algorithm", IEEE 0-7803-8359-1/04/$20.00 0 2004

[15] Canis_familiaris.BAOADD2.66.pep.abinitio.fa.gz.

Passing VBR in Mobile Ad Hoc Networks – for effective live video Streaming

V. Saravanan
Asst.Professor in Computer Applications,
Hindusthan College of arts and Science,
Coimbatore, India

Dr. C.Chandrasekar
Associate Professor,
Dept.of Computer Science,
Periyar University, Salem, India

***Abstract*—Mobile ad hoc networks (often referred to as MANETs) consist of wireless hosts that communicate with each other in the absence of a fixed infrastructure. This technique can be used effectively in disaster management, intellectual conference and also in the battlefield environments. It has the significant attention in the recent years. This research paper depicts the remuneration of using suggestion tracking for selecting energy-conserving routes in delay-tolerant applications and it sends Variable Bit Rate delivery. The previous investigation set up from earlier period surveillance that delay can be traded for energy efficiency in selecting a path. The Prior objective is to find an experiential upper bound on the energy savings by assuming that each node accurately knows or predicts its future path. It examines the effect of varying the amount of future information on routing. Such a bound may prove useful in deciding how far to look in advance, and thus how much convolution to provide in mobility tracking.**

Keywords-Variable Bit Rate; Mobile Ad Hoc; Machine Learning.

I. Introduction

Mobile ad hoc networks are a set of peer-to-peer reconfigurable networks. In general, any pair of associated nodes may communicate with one another, using transitional nodes to store and forward frames.

The annoyance of having a flat hierarchy is that the capacity of the network is expended forwarding other nodes' data. In this paper, we concentrate on Mobile Ad Hoc Networks with a large number of nodes with delay-broadminded applications. Under these conditions, properly delaying forwarding can greatly increase the transport capacity or, equivalently, the lifetime of the network. In addition, it is used in accounting for their complexities. For even greater savings, mobility tracking may be augmented by the following machine learning methods:

- Mechanical classification coordination to avoid channel disputation - passage-tracking
- To keep up connectivity by significant how much energy is absent: Energy-tracking
- To stay away from dead ends with geographic steering: Route Map-structure.

II. Variable Bit Rate

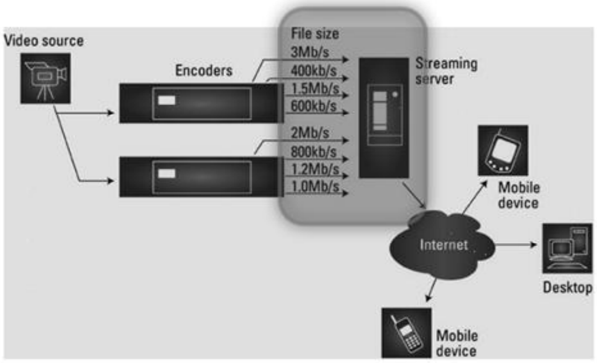


Figure 1. Process VBR in MANET

Variable Bitrate (VBR) is a term used in telecommunications and computing that relates to the bitrate used in sound or video encoding. As opposed to constant bitrate (CBR), VBR files vary the amount of output data per time segment. VBR allows a higher bitrate (and therefore more storage space) to be allocated for more complex segments of media files while less space is allocated to less complex segments. The average of these rates can be calculated to produce an average bitrate for the file.

MP3, WMA, Vorbis, and AAC audio files can optionally be encoded in VBR. Variable bit rate encoding is also commonly used on MPEG-2 video, MPEG-4 Part 2 video (Xvid, DivX, etc.), MPEG-4 Part 10/H.264 video, Theora, Dirac and other video compression formats. Additionally, the variable rate encoding is inherent in lossless compression schemes such as FLAC and Apple Lossless.

VBR produces a better quality-to-space ratio compared to a CBR file of the same data. The bits available are used more flexibly to encode the sound or video data more accurately, with fewer bits used in less demanding passages and more bits used in difficult-to-encode passages.

III. Map Reading

Our node by node classification faced with the task of establishing a route in a MANET, one's first thought might be to apply a position based routing algorithm [3].

While this work can be used in small networks with slowly moving nodes, it is not efficient. When we receive information about the location is available offered by GPS, another option is to use geographic routing [6]. The algorithm's spontaneous appeal has won it status. The focus of the method is simply to forward frames in the direction of the goal. When this is not possible, a contingency plan is employed, often at a steep cost in traffic and delay, or the link is dropped. By mobility tracking we refer to the act of keeping track of a node's position in order to estimate where it will be in the future. In this simple case, which we study here, each node tracks only itself. This general case, each node tracks some other nodes, using a distributed algorithm. The research on mobility tracking in the context of wireless networks has usually focused on dead reckoning [8], only recently considering more sophisticated approaches involving machine learning. Greater attention has been paid to this issue by the robotics community [1] and the pervasive computing community [10].

IV. SCALABILITY ISSUES IN ROUTING MANET

Routing in wireless networks, especially with a large number of nodes, requires a different approach from those in fixed networks. The reason is that conventional routing algorithms attempt to minimize the number of bounds without concern for the overhead incurred by the direction-finding algorithm. This strategy is not successful in mobile ad hoc networks because the routing overhead constitutes a greater fraction of the overall traffic. Therefore, it imposes a fundamental limit on the scalability of wireless networks. Furthermore, different applications may demand optimizing different metrics like delay, dependability, and network lifetime. The rapidly changing topology exacerbates the problem, as the location manager struggles to maintain accurate estimates of the nodes' location. This assumes that reducing the end-to-end delay is a priority. If not, then it is actually possible to increase the transport capacity by using diversity routing.

V. ALGORITHM

To the best of our acquaintance, the closest results are analyzing the impact of delay on throughout in. However, they assume that only one intermediate is forwarded through and that diversity routing and diversity coding are used. In contrast, because of VBR passing, we are interested in characterizing the energy savings by providing future location information, for following belongings:

- Machine learning: Mechanical classification coordination to avoid channel disputation - passage-tracking
- Future is known only one's own.
- The future of one's neighbors is also known.
- To keep up connectivity by significant how much energy is absent: Energy-tracking by Machine Learning?
- To stay away from dead ends with geographic steering: Route Map-structure by Machine learning

Our goal is simply to make the best forwarding decision for geographic routing next bound in order to extend the lifetime of the network. We intend to determine the energy savings of a simulated with respect to the amount of future knowledge and maximum permissible delay. We consider the network lifetime and the end-to-end route power.

A. Machine Learning

Mechanical classification coordination to avoid channel disputation - passage-tracking. In this phase, we classify the traffic tracking by using cameras [7].

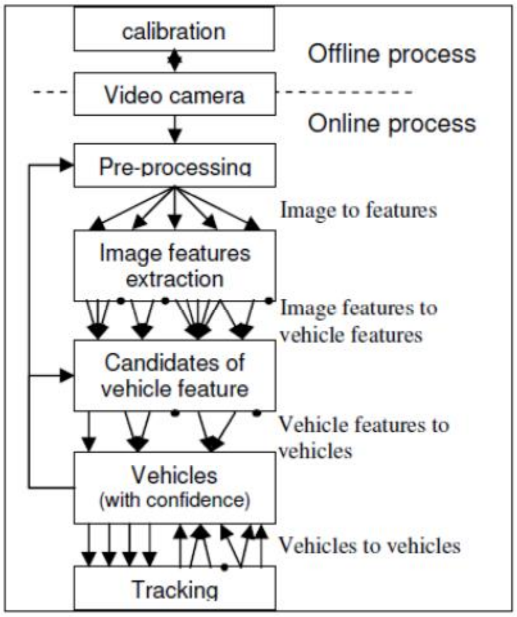

calibration
Offline process
Video camera
Online process
Pre-processing
Image to features
Image features extraction
Image features to vehicle features
Candidates of vehicle feature
Vehicle features to vehicles
Vehicles (with confidence)
Vehicles to vehicles
Tracking

Figure 2. Vehicle Features Detection and Tracking.

It's a process of traffic tracking b using cameras in the mode of machine learning system

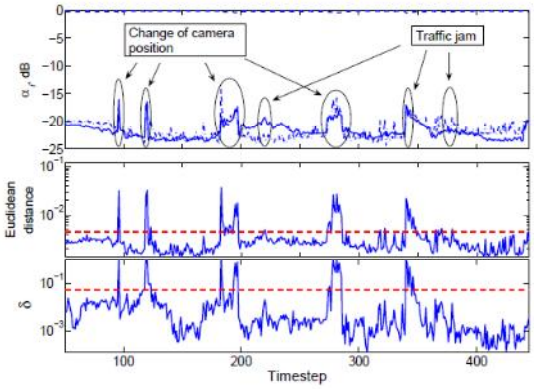


Figure 2.1. The plot examine the results

Traffic jam and time taken in real time traffic ad hoc network by using cameras. Top panel: Annotated plot of average wavelet coefficients in sub bands evolving through time. Approximation coefficients are shown by dashed line. Solid and dashed lines indicate transform levels 1 and 6 corresponding to highest and lowest frequency. Middle panel: OCNM using K th nearest-neighbor distance, with dashed line indicating 90% MVS threshold. Bottom panel: projection error _t with dashed line indicating lower threshold. Transports Quebec dataset [7].

B. Route Map – Structure by machine learning

We use the example in Figure. 3 to illustrate the basic ideas in node bound identification system. We assume that some nodes have already been selected as landmark nodes by the Node place range and each node knows its bound distance to all node location. In Fig. 1, L_1, L_2 and L_3 are three possible locations. Following a predefined order, the bound distance of a node to all the areas is combined into a vector, i.e. the node's bound reference identification number. For example, L_2's bound number is 305 in Figure. 3, representing that L_2 is 3 bounds away from L_1, 0 bound away from itself and 5 bounds away from L_3.

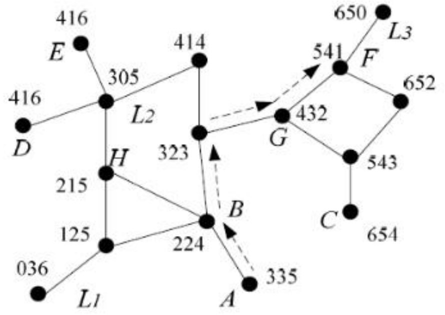


Figure 3. To stay away from dead ends with geographic steering: Route Map-structure.

Figure. 3 Example of Bound number. A node N's Bound Number xyz means N is x, y, z bounds away from landmark L_1, L_2 and L_3 respectively. Intuitively, the Bound Number can reflect the proximity of the network to some extent. Take two nodes N_1 and N_2 for example, we define the bound distance between N_1 and N_2 as L_h. Assume there are m landmark nodes, and the Bound Number of N_1 is

$H_K^{(1)}$ $(H_1^{(1)}, H_2^{(1)}, \ldots H_m^{(1)})$, bound number of N_2 is $H_K^{(2)}$ $(H_1^{(2)}, H_2^{(2)}, \ldots H_m^{(2)})$, the following triangulation inequality holds:

$\mathrm{Max}_K (|H_K^{(1)} - H_k^{(2)}|) \leq L_h \leq \mathrm{Min}_K (H_k^{(1)} + H_k^{(2)})$ (1).

according to the grapevine, for each k from 1 to m, L_h is no more than the sum of $H_K^{(1)}$ and $H_K^{(2)}$, since there exists a path from N_1 to N_2 via landmark k and the bound count of this path is $H_k^{(1)} + H_k^{(2)}$. For the left part of the inequality, without losing the generality, we assume $H_k^{(1)}$ is no more than $H_k^{(2)}$. $H_k^{(2)}$ is no more than the sum of L_h and $H_k^{(1)}$, because there is a path from landmark k to N_2 via node N_1 and $H_k^{(2)}$ is the shortest bound distance from landmark k to $N2$. These inequalities yield a lower bound L and an upper bound U of L_h. More landmark nodes can make the lower and upper bounds. The number of landmarks needed in reality will be a constant which is determined by the precision requirement other than number of nodes in the network. By using above algorithm process, we effectively route make the best forwarding decision for geographic routing in order to extend the lifetime of the network. We intend to determine the energy savings of a simulated MANET with respect to the mount of future knowledge and maximum permissible delay.

VI. MAP READING

Wireless networks of sensors are likely to be widely deployed in the near future because they greatly extend our ability to monitor and control the physical environment from remote locations and improve our accuracy of information obtained via collaboration among sensor nodes and online information processing at those nodes.

Our goal is simply to make the best forwarding decision for geographic routing next bound in order to extend the lifetime of the network. We intend to determine the energy savings of a simulated with respect to the amount of future knowledge and maximum permissible delay. We regard as the network lifetime and the end-to-end route power. In further research is to implement this functionality fully in machine learning system.

ACKNOWLEDGMENT

Our thanks to the experts who have contributed towards development of the work.

REFERENCES

[1] Abdelsalam, W. and Ebrahim, Y., Managing Uncertainty:Modeling Users in Location-Tracking Applications, IEEE Pervasive Computing, vol. 3, no. 3, pp. 60–65,2004

[2] BENNEWITZ, M., BURGARD, W., CIELNIAK, G. AND THRUN, S., Learning Motion Patterns of People for Compliant Robot Motion, The International Journal of Robotics Research, vol. 24, no. 1, pp. 31–48, Jan. 2005 DOI:10.1177/0278364904048962.

[3] GUPTA, P. AND KUMAR, P. R., The Capacity of Wireless Networks, IEEE Transactions on Information Theory, vol. 46, no. 2, pp. 388–404, Mar. 2000 DOI:10.1109/18.825799.

[4] C. Scott and R. Nowak, Learning minimum volume sets, J. Machine Learning Research (JMLR), vol. 7, pp. 665-704, Apr. 2006.

[5] KWOK, C., FOX, D. AND MEILˇA, M., Real-Time Particle Filters, Proceedings of the IEEE, vol. 92, no. 3, pp. 469–484, Mar. 2004 DOI : 10.1109 / JPROC.2003.823144.

[6] STOJMENOVIC, I., Position-Based Routing Algorithms in Ad Hoc Networks, IEEE Com

[7] Machine Learning Approaches to Network Anomaly Detection, Tarem Ahmed, Boris Oreshkin and Mark Coates Department of Electrical and Computer Engineering McGill University Montreal, QC, Canada

[8] KUMAR, V. AND DAS, S. R., Performance of Dead Reckoning-Based Location Service for Mobile Ad Hoc Networks, Wiley Wireless Communications and Mobile Computing, vol. 4, no. 2, pp. 189–202, Mar. 2004 DOI:10.1002/wcm.163.

[9] L. Ruan, H. Du, X. Jia, W. Wu, Y. Li, and K.-I. Ko. A greedy approximation for minimum connected dominating sets. Theoretical Computer Science, 329(1-3):325–330, 2004 unications Magazine, vol. 40, no. 7, pp. 128–134, Jul. 2002 DOI:10.1109/MCOM.2002.1018018.

[10] M. Chatterjee, S. Das and D. Turgut. WCA: A Weighted Cluster- ing Algorithm for Mobile Ad Hoc Networks. Journal of Cluster Computing (Special Issue on Mobile Ad hoc Networks), Vol. 5, No. 2, April 2002, 193-204.

[11] Anindya Iqbal, Nafees Ahmed, and Md. Mostofa Akbar. Directional antenna based connected dominating set construction for energy efficient broadcasting in wireless ad hoc network Computer and Electrical Engineering, International Conference on, 0:839–843,2008.

V.Saravanan received M.Sc(CS)., from Bharathiar University in 1999, completed M.Phil., from Manonmaniam Sundaranar University in 2002. Received MCA., from Periyar University in 2011. Currently working in Asst. Professor in Dept. Of Computer Applications, Hindusthan College of arts and science, Coimbatore. His research area is Network security in mobile networking.

Dr. C.Chandrasekar is an Associate Professor in Dept. Of Computer Science, Periyar University, salem. He has done MCA., Ph.D., He is been in the Teaching and Research field for more than 15 years. His Research area is Mobile and Wireless computing . He has worked and published more than 40 research articles.

Financial Statement Fraud Detection using Text Mining

Rajan Gupta
Research Scholar, Department of Computer Science & Application, Maharshi Dayanand University, Rohtak, Haryana, India

Nasib Singh Gill
Professor, Department of Computer Science & Application, Maharshi Dayanand University, Rohtak, Haryana, India

***Abstract*—Data mining techniques have been used enormously by the researchers' community in detecting financial statement fraud. Most of the research in this direction has used the numbers (quantitative information) i.e. financial ratios present in the financial statements for detecting fraud. There is very little or no research on the analysis of text such as auditor's comments or notes present in published reports. In this study we propose a text mining approach for detecting financial statement fraud by analyzing the hidden clues in the qualitative information (text) present in financial statements.**

Keywords-Text Mining; Bag of words; Support Vector Machines.

I. INTRODUCTION

The illegitimate task of financial statement fraud had considerably affected the economy of a company. The analysis of financial statements assists the capital market participants in deciding about investing in a company. The information present in these statements express the performance of an organization in terms of financial status to the interested parties such as investors, creditors, auditors and management. Any deviation from Generally Accepted Accounting Principles such as presence of some extraordinary values in financial statements may results in a fraud. The presence of deviation does not always results in fraud because departures from GAAP may be appropriate to the company's situation and such departure may have been adequately disclosed.

Detection of financial statement fraud is a difficult task because of the nature of financial statements and warning signs. The mere presence of warning signs does not guarantee the occurrence of fraud and it is difficult to assess their impact before the entire fraud has unraveled. This problem is aggravated further by the fact that financial statements can be misleading even if they are in accordance with GAAP.

Financial statements released by companies consist of textual information in form of auditor's comments and disclosure as footnotes along with financial ratios. This qualitative information may contain indicators of fraudulent financial reporting in form of strategically placed phrases. In order to conceal the fraudulent activity, perpetrators may use selective sentence constructions, selective adjectives and adverbial phrases. Financial statement fraud can be detected by analyzing the above mentioned signals hidden in textual information present in published financial reports.

Companies may present a rosy picture to the investors by manipulating the financial measurements and qualitative narratives of financial statements. These disclosures (qualitative narratives) may not contain fraud indicators explicitly; however indicators of fraud can be constructed by understanding the syntactic as well as semantics of any natural language because perpetrators of fraud may camouflage the indicators by using semantic arsenal of the language. Therefore, in order to detect fraud, it is necessary to examine the qualitative disclosures in the footnotes in the financial statements, as well as the numbers (quantitative information) associated with financial statements.

Quantitative information has been analyzed by number of researchers for detection of fraudulent financial reporting. Therefore, in order to detect fraud indicators present in qualitative contents of financial statements, we present a text mining approach for differentiating between fraud and non – fraud financial statements.

The textual information present in financial statements is unstructured in nature. Text is generally amorphous and therefore must be converted into structured data before applying any predictive data mining techniques such as classification or unsupervised learning method such as clustering in order to detect fraudulent financial reporting.

Text mining is a process of extracting meaningful numeric indices (structured data) from unstructured text. Text mining can analyze words or cluster of words and can be used for determining the relationship with other variables of interest such as fraud or non fraud. Therefore, a text mining approach for detecting fraudulent financial reporting is presented in this paper. The rest of the paper is organized as follows. Section 2 presents a brief overview of the research done in the field of detection of financial statement fraud and identifies the need of an approach for analyzing text present in financial statements for detecting fraudulent financial reporting. Section 3 represents a text mining approach for detection of financial statement fraud followed by conclusion (Section 4).

II. LITERATURE REVIEW

A number of researchers have devoted a significant amount of effort in detecting fraudulent financial reporting. In order to detect fraud several researchers have used various data mining techniques.

For instance, Koh and Low [1] constructed a decision tree by using a data sample of 165 organizations. In order to detect fraud, following six financial variables were examined: quick assets to current liabilities, market value of equity to total assets, total liabilities to total assets, interest payments to earnings before interest and tax, net income to total assets, and retained earnings to total assets. Cecchini M. [2] in 2005 examined quantitative variables along with text information for detection of fraud. The qualitative variables were mapped to a higher dimension which takes in to account ratios and year over year changes.

Kotsiantis et al [3] explored the effectiveness of machine learning techniques such as Decision Tree, Artificial Neural Network, Bayesian Network, K – Nearest Neighbour, Support Vector Machines in detecting firms that issue fraudulent financial statements. The 41 fraudulent firms were matched with 123 non- fraudulent firms. All the variables used in the sample were extracted from formal financial statements, such as balance sheets and income statements.

In 2007, Kirkos et al [4] investigated the usefulness of three Data Mining classification methods namely Decision Trees, Neural Networks and Bayesian Belief Networks by analyzing 27 financial ratios extracted from publicly available data of 76 Greek manufacturing firms for detecting fraudulent financial statements. Further, Hoogs et al [5] developed a genetic algorithm approach for detecting financial statement fraud by analyzing 76 comparative metrics, based on specific financial metrics and ratios that capture company performance.

Belinna et al [6] examined the effectiveness of CART on identification and detection of financial statement fraud by analyzing financial ratios from financial reports of 148 organizations and found CART as a very effective technique in classifying financial statements as fraudulent or non – fraudulent.

Ibrahim et al [7] examined the efficiency of data mining techniques i.e. decision tree and neural network for detection of financial statement fraud by analyzing data from 100 manufacturing firms and concluded that leverage ratio and return on assets ratios are important financial ratios in detecting financial statement fraud.

Furthermore, Ravishankar et al [8] in 2011 applied six data mining techniques namely Multilayer Feed Forward Neural Network (MLFF), Support Vector Machines (SVM), Genetic Programming (GP), Group Method of Data Handling (GMDH), Logistic Regression (LR), and Probabilistic Neural Network (PNN) to identify companies that resort to financial statement fraud on a data set obtained from 202 Chinese companies of which 101 were fraudulent and 101 were non-fraudulent companies. The input vector used by them consists of 35 financial variables or ratios extracted from publically available financial statements.

Recently, Gupta et al [9] examined the efficacy of three data mining techniques namely CART, Naïve Bayesian Classifier and Genetic Programming for detecting financial statement fraud by analyzing 52 financial ratios extracted from financial statements of 114 organizations.

The review of existing academic literature reveals that research conducted till date in the field of detection of financial statement fraud had majorly analyzed financial ratios or variables which can be extracted from financial statements. A very few studies have analyzed the key component of financial statements i.e. qualitative contents in order to detect fraud.

In order to detect hidden valuable knowledge from textual financial data, we propose a text mining approach in this study because traditional mining techniques are insufficient in detecting fraud from the increasing amount of text data.

III. Text Mining: An Approach For Detection Of Financial Statement Fraud

Figure 1 illustrates the proposed text mining approach for financial statement fraud detection. Text mining system takes as an input the collection of financial statements. In order to detect fraudulent financial reporting, financial statements of both type of organizations (fraudulent or non fraudulent) need to be collected as the first step. Companies with fraudulent history can be identified by analyzing AAER's issued by SEC. Data set should contain financial statements of non fraud organization for each fraudulent organization. The non fraud organization should be of same size (on the basis of assets or sales) as that of fraudulent organizations.

Second step is preprocessing which involves the extraction of qualitative narratives from financial statements and arranging into a document because a document is a basic unit of analysis in text mining. During preprocessing, words present in all the documents should be converted into lower case so as to avoid inclusion of two same words such as "Legal" and "legal" as different words in the corpus (collection of documents).

All the punctuations should be removed from the corpus followed by removal of any number if present because input to the classifiers should contain only text. Stopwords such as articles (a, the etc.), conjunctions (but, and etc.) and prepositions (on, in etc.) should also be removed during preprocessing because these words does not help in discriminating the documents. Stemming is not required in domain of accounts because inflected terms may have different meanings.

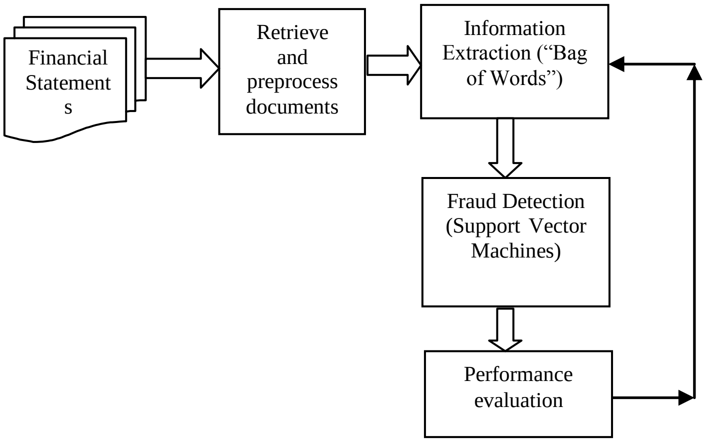

Financial Statement s
Retrieve and preprocess documents
Information Extraction ("Bag of Words")
Fraud Detection (Support Vector Machines)
Performance evaluation

Figure 1: Text Mining detection for financial statement fraud

Since, in text mining, a sentence is regarded as a set of words and order of words can be changed with no impact on the result of the analysis, therefore syntactical structure of a sentence can be ignored for handling the text in an efficient manner. However, information regarding number of occurrences of each word should be retained. This unordered collection of words is known as "bag of words". In "bag of words" approach, the occurrence of each word is used as a feature for training a classifier. The "bag of words" model represents each document with a vector of word count that appears in the document. The vector associated with each document is compared with typical vector associated with a given class (fraud or non fraud). Documents with similar vectors are considered to be similar in content and dissimilar otherwise.

The vector spaces generated above will be used by next step for classifying organizations into fraud or non fraud. We recommend the use of Support Vector Machine – a supervised classification method, for detecting fraudulent financial reporting because SVM's construct a hyperplane in feature space which best classifies among fraudulent or non fraudulent financial reporting. SVM takes a set of input data and predicts, for each given input, which of two possible classes (fraud or non fraud) forms the output. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other.

Since SVM is a supervised machine learning method, it will learn from feature spaces of both fraudulent and non fraudulent examples present in the training set. After learning, this method is capable of classifying correctly between fraud and non fraud organizations present in the testing dataset. The accurateness of classification should be evaluated by using evaluation measures such as accuracy, precision, recall (sensitivity in binary classification), F-measure and purity.

IV. CONCLUSION

In this conceptual paper, we presented a text mining approach for detection of financial statement fraud. Fraud detection model presented in this paper begins with collection of financial statements for both fraud and non fraud organizations followed by preprocessing which involves lexical analysis of text present in financial statements. At the next step, bag of words approach has been selected for extracting information hidden in the text which results in vector spaces for both fraudulent and non fraudulent organizations.

These vector spaces acts as an input vector to the Support vector machines which learns from training data and further classifies organizations from testing data into fraud or non fraud. Finally, the correctness of classification is measured by using standard evaluation measures.

The methodology proposed in this paper for detection of financial statement fraud differs from earlier methodology in terms of input vector. Input vector in most of the previous studies consists of financial ratios and metrics i.e. quantitative information present in financial statements. Unlike earlier research studies, we selected text i.e. qualitative narratives present in financial statements in order to assess likelihood of financial statement fraud.

Financial statement fraud is a major concern for most of the organization worldwide. Hence both the quantitative and qualitative information available in annual reports should be analyzed simultaneously for assessing the risk of fraud.

REFERENCES

[1] H.C. Koh, C.K. Low, Going concern prediction using data mining techniques, Managerial Auditing Journal 19 (3) (2004) 462–476.

[2] Cecchini M. 2005. Quantifying the risk of financial events using kernel methods and information retrieval. Doctoral dissertation, University of Florida.

[3] Kotsiantis S., Koumanakos E., Tzelepis D. and Tampakas V. "Forecasting Fraudulent Financial Statements using Data Mining", International Journal of Computational Intelligence VOLUME 3 NUMBER 2 2006.

[4] Efstathios Kirkos, Charalambos Spathis &Yannis Manolopoulos (2007), Data mining techniques for the detection of fraudulent financial statements. Expert Systems with Applications 32 (23) (2007) 995–1003.

[5] Hoogs Bethany, Thomas Kiehl, Christina Lacomb and DenizSenturk (2007). A Genetic Algorithm Approach to Detecting Temporal Patterns Indicative Of Financial Statement Fraud, Intelligent systems in accounting finance and management 2007; 15: 41 – 56, John Wiley & Sons, USA, available at: www.interscience.wiley.com.

[6] BelinnaBai, Jerome yen, Xiaoguang Yang, False Financial Statements: Characteristics of china listed companies and CART Detection Approach, International Journal of Information Technology and Decision Making , Vol. 7, No. 2(2008), 339 – 359.

[7] Ibrahim H. , Ali H. "The use of data mining techniques in detecting fraudulent financial statements: An application on manufacturing firms", The journal of faculty of economics and administrative sciences, (2009) Vol. 14, No. 2 pp. 157 – 170.

[8] P.Ravisankar, V. Ravi, G.RaghavaRao, I., Bose, Detection of financial statement fraud and feature selection using data mining techniques, Decision Support Systems, 50(2011) 491 – 500.

[9] Gupta Rajan, Gill N.S. 2012 "Data Mining Techniques – A Key for detection of financial statement fraud" , International Journal of Computer Science and Information Security, Volume 10 No. 3, pp. 49 – 57.

Error Analysis on Estimation Method for Air-Temperature, Atmopspheric Pressure, and Realtive Humidity Using Absorption Due to CO_2, O_2, and H_2O Which situated at Around Near Infrared Wavelength Region

Kohei Arai [1]
Graduate School of Science and Engineering
Saga University
Saga City, Japan

***Abstract*—A method for air-temperature, atmospheric pressure and relative humidity using absorptions due to CO_2, O_2 and H_2O which situated at around near infrared wavelength region is proposed and is evaluated its validity. Simulation study results with MODTRAN show a validity of the proposed method.**

Keywords-absorption band; regressive analysis; air-temperature; atmospheric pressure and relative humidity estimations.

I. Introduction

Hyperspectrometer in the visible to near infrared wavelength regions are developed and used for general purposes of earth observation missions such as Agriculture, Mineralogy, Surveillance, Physics, Chemical Imaging, Environment, in particular, for mineral resources explorations and agricultural monitoring [1]-[15]. Hyperspectormeter allows estimate atmospheric continuants by using absorption characteristics of the atmospheric continuants because spectral bandwidth of the hyperspectrometer is quit narrow like an atmospheric sounders onboard earth observation satellites [16].

The aim of the paper is to propose the method for estimation of air-temperature, atmospheric pressure, and relative humidity on the sea level together with estimation accuracy assessment with the different bandwidth. Method for air-temperature, water vapor and atmospheric pressure estimations with spectral radiometer in near infrared wavelength regions is proposed. It can be assumed that there is no up-welling radiance from the ocean in near infrared wavelength regions. Therefore, the major contribution of the observed radiance is assumed to be derived from the atmosphere. Thus it is possible to estimate atmospheric continuants, oxygen, carbon dioxide, water vapor concentrations can be estimated.

There are absorption bands due to O_2, CO_2 and H_2O in the near infrared wavelength regions, 762nm, 1382nm and 980nm, respectively. It is possible to estimate atmospheric pressure (O_2), air-temperature (CO_2) and relative humidity (H_2O) by measuring the ocean at the wavelength of 762, 980 and 1382nm, respectively. Therefore, atmospheric pressure, air-temperature, and relative humidity is estimated.

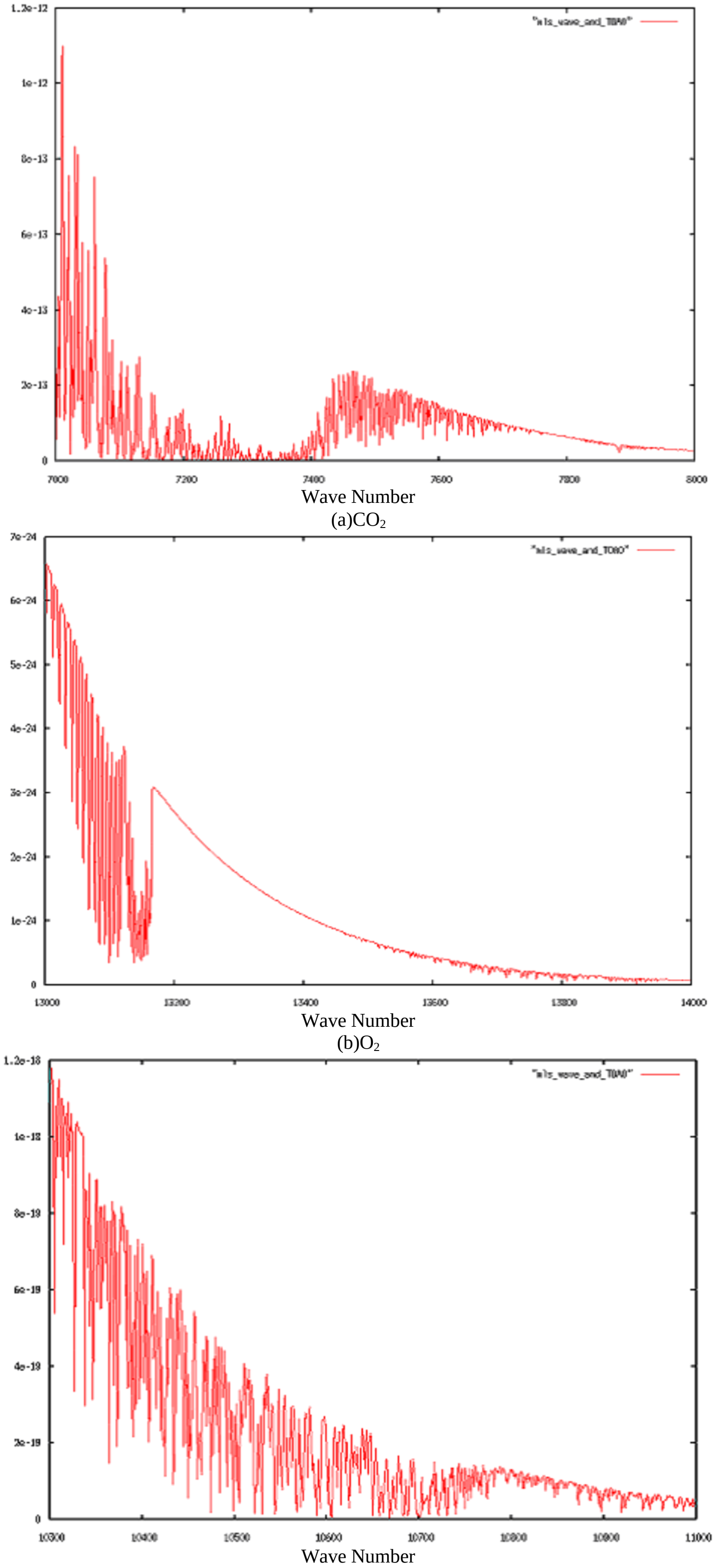
By using MODTRAN, the Top of the Atmosphere: TOA radiance, or at sensor radiance is calculated at the aforementioned wavelength with the different band width, 1, 2, 4, 8nm. TOA radiance for 10 bands, 5 bands, 2 bands and 1 band are calculated for 1, 2, 4, 8nm bandwidth at around 762, 980 and 1382nm. By using TOA radiance, regressive analysis is made based on logarithmic function. Regressive coefficients and Root Mean Square Error: RMSE are calculated for accuracy assessment.

The following section describes the proposed method for estimation of air-temperature, atmospheric pressure, and relative humidity with hyperspectrometer data followed by simulation study for assessment of estimation accuracy. Then conclusion is followed together with some discussions.

II. Proposed Method

A. *Absorption Characteristics of Oxygen, Carbon Dioxide, and Water Vapor*

Figure 1 shows absorption characteristics of oxygen, carbon dioxide, and water vapor in the near infrared wavelength regions, 1394 to 1406 nm, 756 to 768 nm, and 934 to 946 nm which are corresponding to 7000 to 8000 cm^{-1}, 13000 to 14000 cm^{-1}, and 10300 to 11000 cm^{-1}, respectively. The vertical axis of Figure 1 shows TOA radiance in unit of $W/m^2/str$. As mentioned above, it is possible to estimate oxygen (Atmospheric Pressure), carbon dioxide (Air-Temperature), and water vapor (Relative Humidity) concentrations by using these absorption characteristics.



(a)CO_2

(b)O_2

(c)H_2O

Figure 1. Absorption characteristics of CO2, O2, H2O for estimation of air-temperature, atmospheric pressure and relative humidity with hyper-spectrometer

B. Procedure of the Proposed Method for Estimation of Air-Temperature, Atmospheric Pressure, and Relative Humidity

Assuming up-welling radiance from the ocean is negligible, at sensor radiance of hyperspectrometer is reflected by the absorption characteristics of oxygen, carbon dioxide, and water vapor. Using the at sensor radiance, Atmospheric Pressure: AP, Air-Temperature: AT, and Relative Humidity: RH is estimated with the following regressive equations.

$$AT = a_0 + a_1 ln(b_1 TOA_1) + a_2 ln(b_2 TOA_2) + \_\ \_\ \_ + a_n ln(b_n TOA_n) \qquad (1)$$

$$AP = \mathrm{c}_0 + \mathrm{c}_1 ln(d_1 TOA_1) + \mathrm{c}_2 ln(d_2 TOA_2) + \_\ \_\ \_ + \mathrm{c}_n ln(d_n TOA_n) \qquad (2)$$

$$RH = \mathrm{e}_0 + \mathrm{e}_1 ln(f_1 TOA_1) + \mathrm{e}_2 ln(f_2 TOA_2) + \_\ \_\ \_ + \mathrm{e}_n ln(f_n TOA_n) \qquad (3)$$

where TOAn denotes at sensor radiance for band number n while a to f denotes regressive coefficients. In these equations, Beer-Bouque-Lambert law is assumed for radiative transfer processes in the atmosphere.

III. Experiemnts (Simulation Studies)

A. Simulation Data Used

Utilizing MODTRAN of radiative transfer code, at sensor radiance is calculated by wave number by wave number. Bandwidth can be changed in the calculation of at sensor radiance. Other atmospheric conditions are set at the default values of Mid. Latitude Summer of atmospheric model which are included in the MODTRAN.

Air-Temperature, Atmospheric Pressure, and Relative Humidity are set at the default values and the default value plus minus 30% of additive biases as shown in Table 1. At sensor radiance is calculated with MODTRAN.

B. Simulation Results

Using these calculated at sensor radiance, regressive analysis is conducted based on the regressive equations, equation (1) to (3). Through the regressive analysis, regressive coefficients are determined together with Root Mean Square Error: RMSE, regressive error. Table 2 shows the results from the regressive analysis. Bandwidth are set at 1, 2, 4, and 8 nm which are reasonable ranges from the state of the art on hgyperspectrometer design and development.

As shown in Table 2, the regressive errors for Air-Temperature, Atmospheric Pressure, and Relative Humidity range from 0.033 to 1.61 (%), from 0.59 to 1.06 (%), and from 0.096 to 1.28 (%), respectively.

Although it is supposed that the regressive error of the 1nm bandwidth case is the best followed by the 2nm bandwidth case, and so on for all the geophysical parameters, Air-Temperature, Atmospheric Pressure, and Relative Humidity, it is no always true. For instance, the regressive error of the 2 nm bandwidth case is smaller than that of the 1 nm bandwidth case for atmospheric pressure..

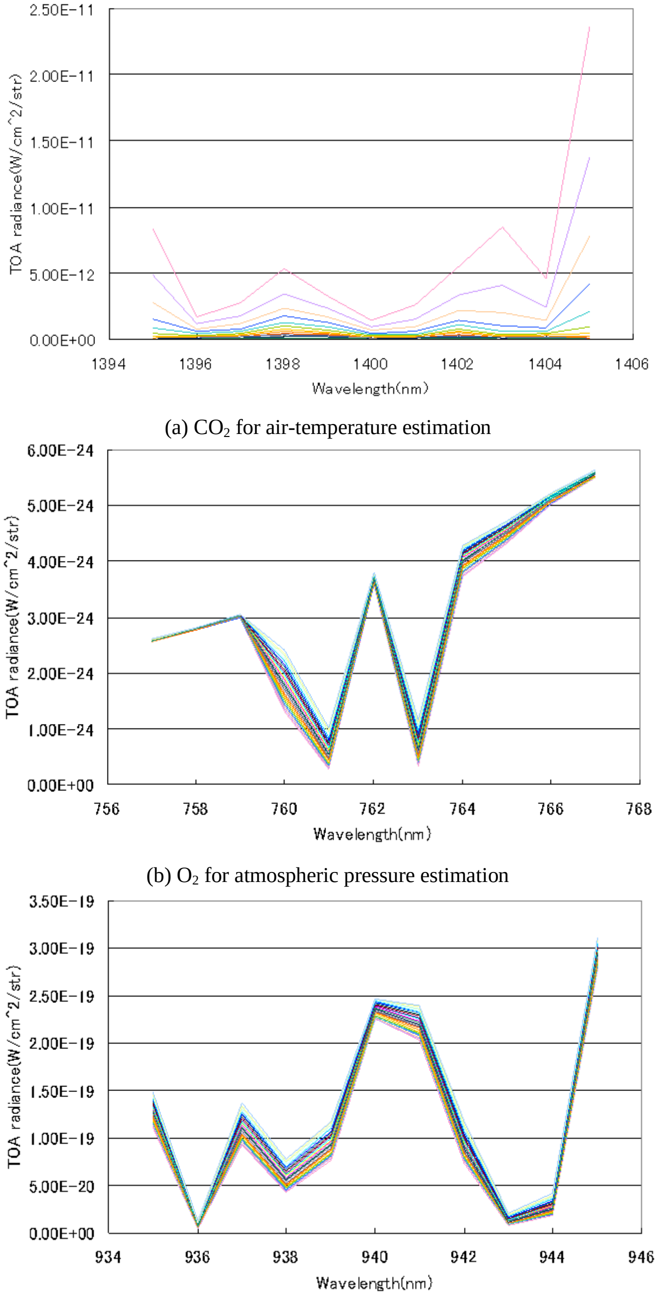
TABLE I. Parameters Set To MODTRAN For TOA Radiance Calculations

Additive Bias(%)	Air-Temp.[K]	Atm.Press[hPa]	RH[%}
30	382.46	1316.9	99.06
27.5	375.105	1291.575	97.155
25	367.75	1266.25	95.25
22.5	360.395	1240.925	93.345
20	353.04	1215.6	91.44
17.5	345.685	1190.275	89.535
15	338.33	1164.95	87.63
12.5	330.975	1139.625	85.725
10	323.62	1114.3	83.82
7.5	316.265	1088.975	81.915
5	308.91	1063.65	80.01
2.5	301.555	1038.325	78.105
0	294.2	1013	76.2
-2.5	286.845	987.675	74.295
-5	279.49	962.35	72.39
-7.5	272.135	937.025	70.485
-10	264.78	911.7	68.58
-12.5	257.425	886.375	66.675
-15	250.07	861.05	64.77
-17.5	242.715	835.725	62.865
-20	235.36	810.4	60.96
-22.5	228.005	785.075	59.055
-25	220.65	759.75	57.15
-27.5	213.295	734.425	55.245
-30	205.94	709.1	53.34

TABLE II. Regressive Error For Estimation Of Air-Temperature, Atmospheric Pressure And Relative Humidity With Hyper-Spectrometer Data

	Bandwidth(nm)	Default	Estimated	Difference
Air-Temperature	1	294.2	294.298	0.098
	2	294.2	294.445	0.245
	4	294.2	294.935	0.735
	8	294.2	299.05	4.85
Atm.Pressure	1	1013	1005.77	-7.23
	2	1013	1007.02	-5.98
	4	1013	1002.2	-10.8
	8	1013	1003	-10
Rel.Humidity	1	76.2	76.1886	-0.0114
	2	76.2	76.1556	-0.0444
	4	76.2	76.19266	-0.00734
	8	76.2	75.228	-0.972

Also the regressive error of the 8 nm bandwidth case is better than that of the 4 nm bandwidth case for atmospheric pressure while the regressive error of the 4 nm bandwidth case is smaller than that of the 2 nm bandwidth case. This is because that the wavelength at which absorption starts is different among the bandwidth cases as shown in Figure 2. Figure 2 shows absorption characteristics of oxygen, carbon dioxide, and water vapor with 1 nm interval. Therefore, the band, or spectral response and bandwidth have to be determined properly by referring to the absorption characteristics. Otherwise, it is impossible to determine the best bandwidth.



(a) CO_2 for air-temperature estimation

(b) O_2 for atmospheric pressure estimation

(c)H_2O for relative humidity estimation

Absorption characteristics of CO2, O2, H2O for estimation of air-temperature, atmospheric pressure and relative humidity with hyper-spectrometer

IV. Conclusion

Method for air-temperature, atmospheric pressure and relative humidity using absorptions due to CO_2, O_2 and H_2O which situated at around near infrared wavelength region is proposed and is evaluated its validity. Simulation study results with MODTRAN show a validity of the proposed method.

It is found that the regressive errors for Air-Temperature, Atmospheric Pressure, and Relative Humidity range from 0.033 to 1.61 (%), from 0.59 to 1.06 (%), and from 0.096 to 1.28 (%), respectively. Also it is not always true that narrowest bandwidth shows the best estimation accuracy. Spectral responses of hyperspectrometer would be better to determine by referring absorption characteristics precisely.

Acknowledgment

The author would like to thank Mr. Ken Yamaguchi for his effort to conduct simulation studies.

References

[1] Lushalan Liao, Peter Jarecke, “Radiometric Performance Characterization of the Hyperion Imaging Spectrometer Instrument”, Proc. Optical Science and Technology Symposium, Earth Observing Systems V, SPIE 1435, (2000)

[2] Peter Jarecke, Karen Yokoyama, “Radiometric Calibration Transfer Chain from Primary Standards to the End-to-End Hyperion Sensor”, Proc. Optical Science and Technology Symposium, Earth Observing Systems V, SPIE 1435, (2000).

[3] Schurmer, J.H., Air Force Research Laboratories Technology Horizons, (2003)

[4] Ellis, J., Searching for oil seeps and oil-impacted soil with hyperspectral imagery, Earth Observation Magazine (2001).

[5] Smith, R.B. Introduction to hyperspectral imaging with TMIPS, MicroImages Tutorial Web site, (Accessed on July 14, 2012),

[6] Lacar, F.M., et al., Use of hyperspectral imagery for mapping grape varieties in the Barossa Valley, South Australia , Geoscience and remote sensing symposium (IGARSS'01) - IEEE 2001 International, vol.6 2875-2877p. doi:10.1109/IGARSS.2001.978191, (2001)

[7] Tilling, A.K., et al., Remote sensing to detect nitrogen and water stress in wheat, The Australian Society of Agronomy, (2006)

[8] Fernández Pierna, J.A., et al., 'Combination of Support Vector Machines (SVM) and Near Infrared (NIR) imaging spectroscopy for the detection of meat and bone meat (MBM) in compound feeds' Journal of Chemometrics 18 341-349 (2004)

[9] Holma, H., Thermische Hyperspektralbildgebung im langwelligen Infrarot, Photonik, (2011),

[10] Werff H. Knowledge based remote sensing of complex objects: recognition of spectral and spatial patterns resulting from natural hydrocarbon seepages, Utrecht University, ITC Dissertation 131, 138p. ISBN 90-6164-238-8 (2006),

[11] Noomen, M.F. Hyperspectral reflectance of vegetation affected by underground hydrocarbon gas seepage, Enschede, ITC 151p. ISBN 978-90-8504-671-4 (2007),.

[12] M. Chamberland, V. Farley, A. Vallières, L. Belhumeur, A. Villemaire, J. Giroux et J. Legault, "High-Performance Field-Portable Imaging Radiometric Spectrometer Technology For Hyperspectral imaging Applications," Proc. SPIE 5994, 59940N, September 2005.

[13] Farley, V., Chamberland, M., Lagueux, P., et al., "Chemical agent detection and identification with a hyperspectral imaging infrared sensor," Proceedings of SPIE Vol. 6661, 66610L (2007).

[14] Kevin C. Gross, Kenneth C Bradley and Glen P. Perram, "Remote indentification and quantification of industrial smokestack effluents via imaging Fourier-transform spectroscopy," Environmental Sci Tech, 44, 9390-9397, 2010.

[15] Tremblay, P., Savary, S., Rolland, M., et al., "Standoff gas identification and quantification from turbulent stack plumes with an imaging Fourier-transform spectrometer," Proceedings of SPIE Vol. 7673, 76730H (2010).

[16] K.Arai, K. Yamaguchi,Atmospheric correction through estimation of atmospheric optical properties with hyperspectrometer data, Proceedings of the 46th General Assembly of Japan Society of Remote Sensing, 22, 2009.

AUTHORS PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission “A” of ICSU/COSPAR since 2008. He wrote 30 books and published 322 journal papers.

Productisation of Service: A Case Study

Nilanjan Chattopadhyay
Institute of Management Technology
Ghaziabad, India

***Abstract*—This paper discusses the issue of Productisation of service, i.e. development of systemic, scalable and replicable service offerings, as implemented by a multinational Consulting organization, engaged in the business of outsourcing and consulting solutions, from their office in India. The literature is quite rich with discussions and debates related to products and services individually, but there seems to be an important deficiency in terms of 'integration' between product design and service elements for supporting new service-product system. In today's flat world the geographic boundaries are getting diminished when firms are expanding seamlessly across the globe. This seamless expansion of the electronic data processing market makes use of the outsourcing as one of its main way to expand into various geographies. Data Sanitization is one of the most sought after service offering made by a consulting firm to protect the sensitive client data from any misuse. The paper attempts to document the process followed by a firm to productize its data sanitization service offering. This documentation will not only help in integration of product and service parameters, but also will be extremely helpful for the organizations worldwide offering service as business.**

Keywords-Services management; productisation; service product; data sanitisation; data security; productised service.

I. INTRODUCTION

In 2005, Kuczmarski & Johnston [1] claimed that one emerging trend in service development that continues to grow is that blending of product and services into a "full experiences", rather than standing as individual entities. Following this trend, Sheth et al [2] described the product-servicisation movement which means a marketable combination of products and services is becoming increasingly important to academics and practitioners. Valminen K. & Toivonen M [3] suggested that the service-Productisation approach which provides the service more or less 'product like' can stimulate the service company to produce new innovation. However, the reverse trend of productisation of established service business has been adopted by a few of the multinational service organisations. This paper documents the issues of development of systemic, scalable and replicable service offerings, as implemented by a multinational Consulting organisation, engaged in the business of outsourcing and consulting solutions, from their office in India

This paper is structured into four sections. The first section surveys the literature to conclude that body of knowledge is rich in discussing about products and services on stand-alone basis, but learning from specific cases of Productisation of service are not well documented. This paper attempts to fill in this void. The second section discusses about the service offering in hand for discussion in this paper, viz. Data sanitization. The third section goes into details of processes followed by a multinational service organization in productizing data sanitization, one of their important service offerings. Lastly the fourth section draws a summary and concludes the discussion.

II. PRODUCTISATION OF SERVICE

The literature is quite rich with discussions and debates related to products and services individually, but there seems to be an important deficiency in terms of 'integration' between product design and service elements for supporting new service-product system.

A systematic development of services is becoming increasingly important when the improvement of companies' competitiveness is pursued. Earlier research by de Brentani [4] has shown that proficiency and effectiveness in new service development contributes significantly to the success of the offering. However, the traditional product development models created for industrial production do not fit as such in services due to the specificities of this part of the economy. Grönroos [5] and later Sundbo and Gallouj [6] included immateriality, process nature and the co-production with the client in the first place as the specificities of services.

Jaakkola et al [7] defined productisation as one possible tool to systematize both the development and the production of services so that continuous innovation, cost efficiency and customer orientation become a part of everyday life. There is not one commonly accepted definition for the Productisation of services. Usually the term refers to making the service offering more or less 'product like', i.e. defining the core process and its outcome so that they become more 'stabile' and visible. Individual needs of customers may be taken into account as small variations in the core service, or through modularization. In the latter practice, customization is achieved through different combinations of modules, each component being provided in a systematic manner. Besides the service elements that are visible to the customer, Edvardsson [8] and Vaattovaara [9] opined that Productisation may concern the service company's internal processes.

Productisation can be restricted to the more accurate defining of already existing services, but more commonly the term includes also some renewal of the service. Because of this, Gallouj and Weinstein [10] described Productisation as a factor that stimulates the service company to produce new innovations. In the present paper, we use the broad view of Productisation, which covers both new and existing services. We focus on the systematization of the service, but include in our perspective the even more advanced practice - modularization.

Service companies attempt Productisation of service for improving competitiveness and performance. Defining, systematizing and concretizing a service make its production more profitable and efficient. When the production process is well-defined, the quality of the service becomes more stable. In addition, the possibilities to accumulate knowledge systematically are improved. Productisation often intensifies the transfer of knowledge and enables the division of work. Finally, Productisation makes the pricing of the service easier. Sipilä [11] suggested that companies may even switch from selling experts' time to selling value propositions with a fixed fee.

All these impacts lead not only to better competitiveness, but they also open possibilities for better management. The producer knows better what he is selling and the customer knows better what he is purchasing.

Thus, the customers also benefit from Productisation. It becomes possible for them to compare the outcome of the service with the service promise and to compare the benefit received with the price of the service. In other words, Productisation facilitates the evaluation of the service. The increased tangibility and concreteness - a characteristic which Edvardsson [8] called 'explicitness' - makes the service more tempting and easier to buy.

The focus of Productisation varies. It can be just a minor change of style or appearance in the service, but it can also mean upgrading of the existing service. Further, the idea may be to extend the company's service portfolio in current markets, or as Jaakkola [7] suggested, to develop a new service to an existing customer need or a totally new service to a new customer need.

Each Productisation process is different depending on the company's aims as well as its strategy. Jaakkola et. al [7] stress that companies should plan and carry out their service development project on their own basis and starting from their own needs. According to Jaakkola et al., the Productisation process consists of seven different stages:

1) assessing the clients' needs and the ways in which they are answered; 2) defining the structure, contents and process of the service; 3) specifying the degree of standardization; 4) concretizing the service (service description, brochures etc.), 5) selecting the principles of pricing; 6) following-up and measuring the success of the service; 7) and anticipating the needs for continuous development. Sipilä [11] has emphasized marketing and piloting as additional stages that should be included in a Productisation process".

III. Data Sanitization, The Service Offering

Data Sanitization is the process of camouflaging sensitive information by overwriting it with realistic looking but false data of a similar type. This process is done deliberately, permanently and irreversibly removing the sensitivity of the data stored in organization, to avoid the data theft. It is also called as data masking.

Data sanitization is usually performed on the certain business critical set of attributes by applying various 'Sanitization Techniques' on those attributes and modifying the values of that attribute from the original value but still maintaining the meaning of the attribute and modified value to be in realistic and valid range of data values. Previous research by Zhong et al [12] found that data sanitization is usually needed for personal and identifiable attributes of the business which has more exposure and needs to be protected in this world of highly globalization.

Data sanitization is typically done in the development and test systems where production data is used as a sample for developing or testing product life cycle including

- Development of a new product
- Enhancements to the existing product or suit of products to make them more suitable for the target customers
- Adding new features to the product lines or testing compatibility of the new features with the existing features of the product.
- Launching promotional offers or knowing the effect of promotional offers onto the business lines
- Development of a system to bid for the product build or buy strategy for a company
- Compliance to regulations in case of the pharmaceutical and health insurance industry

The scope of data sanitization within the organization is not only limited to above mentioned operations, but it is also needed for any information technology projects within the company where non-authorized person (within the company or outside the company) may need to use the production data for any internal development or testing related operations. The data sanitization can be achieved in multiple ways. Depending upon the need of an organization and objective of data sanitization, Carr et al [13] suggested that one should define the most appropriate way of achieving data sanitization and thus obtaining the level of data security that organization needs.

The building block of the data sanitization process contains four main components

- Sensitive data from the organization
- Set of sanitization rules
- Encryption Algorithms
- Sanitized data as output

The basic data sanitization conceptual diagram is as shown in figure 1. It contains components like Sensitive data as defined by the organization or organizational business users, sanitization rules engine that consist of various sanitization rules, encryption algorithms as build by organization.

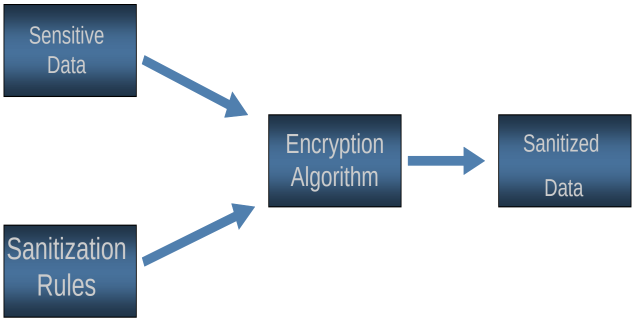


Figure 1.

Sensitive data: Depending upon the nature of the Organization, business users from the organization needs to define the most business critical business elements within the organization. These elements are identified because they are too sensitive for others to see and interpret for any gains. The data may include personal information, financial information such as account numbers, medication information etc. While selecting the sensitive data of an organization business users also needs to take care of the regulatory mandates of various countries where organization does business.

Sanitization Rules: Sanitization rules are defined as set of keys (generic or customized) for each identified sensitive data elements as shown in the figure 1. These rules are usually stored in a form of rule name and numeric rule value parameters. These parameters are fed to the encryption algorithm for rule application. The database of sanitation rules is maintained by an organization so any rule can be applied to any specific or generic field of an organization. This is used to add the strength of the data sanitization.

Encryption Algorithm: The encryption algorithms use the sanitization rules parameters and accordingly select the encryption function to be used. The encryption algorithm with parameters passed by sanitization rules are then applied over the sensitive data for data sanitization.

Sanitized data: The sanitized data is an end product of the sanitization function with completely masked data and helps protect the confidentiality of the data while maintaining the referential integrity and data dependence among attributes. The sanitized data can be used for sharing with other parts of the organization.

The primary need of the data sanitization is to

- Protect the valuable business information to avoid the impact of the data theft on to the business.
- Compliance to regulatory requirement: In almost all countries in the world the regulatory body has defined set of rules regarding the data security. All organizations are subject to comply with the data security. Shringarpure [14] identified some of the regulations as are
- GLB Act: The Gramm-Leach-Bliley Act requires institutions to protect the confidentiality and integrity of personal consumer information.
- Financial Privacy Act of 1978 creates statutory Fourth Amendment protection for financial records and there are a host of individual state laws.
- HIPAA Act 1996: There are also a number of security and privacy requirements for personal information included in the Health Insurance Portability and Accountability Act of 1996
- Sarbanes Oxley Act section 404 requires mandatory security and encryption of the data.
- Enhance the data privacy of an organizational data. As suggested by Zohang [12] data sanitization will help the organization to ensure that the organizational data can be kept safe from any internal breaches. This will ensure that the data does not fall into the wrong hands or data is not exposed to any un-authorized person.

The Gartner Analyst study 2008 shows that

- Approximately 95% of the internal security breaches are avoidable if the necessary proactive steps are taken by the organization.
- More than 50% of all data thefts come from inside the organization from disgruntle employees, employee mistakes, oversights and vendors having improper access.

Prevention cost if the data security is significantly less than the potential losses or the legal mitigation costs

IV. Productisation Of Data Sanitisation Service

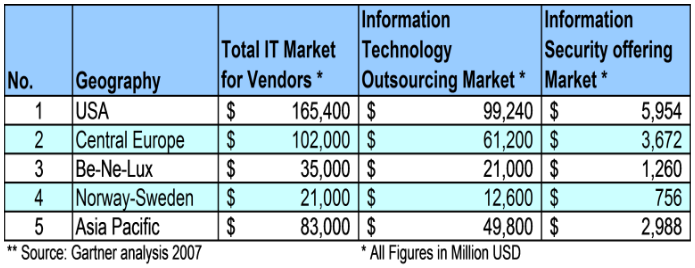
To identify a data sanitization productized service offering requirement for a customer, a consulting firm needs to know the data management plans of an organization. This data management plan of an organization needs to be mapped to regulations, recent security breaches, and best data security practices as followed in the industry where the customer's organization belongs to and last but not the least is an Information Technology vision of an organization [Carr, 13].

V. Market Analysis And Feasibility Study

In today's flat world the geographic boundaries are getting diminished when firms are expanding seamlessly across the globe. This seamless expansion of the electronic data processing market makes use of the outsourcing as one of its main way to expand into various geographies. According to the analyst reports, there are main 5 markets they are USA, Central Europe, Be-Ne-Lux (Belgium, Netherlands and Luxemburg), Norway-Sweden and APAC market. The market for the electronic data processing is growing with many organizations from west (USA and Europe) considering utilizing APAC region to be their main back-office.

As per the Gartner study, it is expected that 6% of the outsourcing market would be a market for Information security spending by an organization. The data security is expected to be around 35% of the total information security market

TABLE I. [16]



No.	Geography	Total IT Market for Vendors *		Information Technology Outsourcing Market *		Information Security offering Market *	
1	USA	$	165,400	$	99,240	$	5,954
2	Central Europe	$	102,000	$	61,200	$	3,672
3	Be-Ne-Lux	$	35,000	$	21,000	$	1,260
4	Norway-Sweden	$	21,000	$	12,600	$	756
5	Asia Pacific	$	83,000	$	49,800	$	2,988

** Source: Gartner analysis 2007 * All Figures in Million USD

The Global Analyst Relations team was formed consist of representation from management, Sales, Development and regional marketing. This team focused on:

Market Analyst and build relations: Drive sales into the identified geographies by influencing perception of market analyst of our capabilities.

Market Knowledge Management: inform internal people about the analysts' opinions of the firm and with in each region as well as globally.

For doing business and making sales of the data sanitization product offering, local knowledge is very important. The localized knowledge was important to define the geography specific business sensitive elements and the regulations specific to the same.

- The broad market segments were defined based on following parameters
- A big geographic presence.
- A market segment has more than $100 Million for data security.
- A segment represents large outsourcing presence i.e. more than USD $10 billion.
- Analyst recommendations for potential data sanitization offer requirements
- Availability of dedicated sales team to take this offer into the market.

Based on the criterion mentioned above, the broad market segments for the Data Sanitization were identified to be

- North America:
- Central Europe:
- Be-Ne-Lux (Belgium, Netherlands and Luxemburg):
- Norway and Sweden:
- APAC (Asia – Pacific)

In all the identified markets, the competitive landscapes were drawn up

VI. The Productized Service Offering

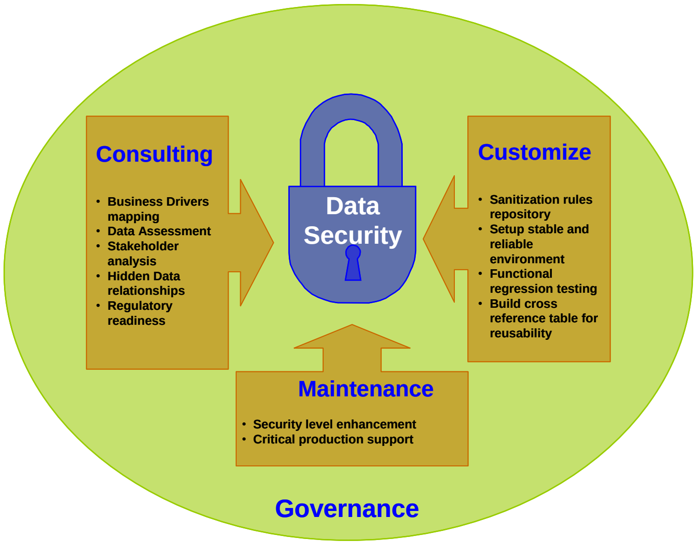

Consulting
• Business Drivers mapping
• Data Assessment
• Stakeholder analysis
• Hidden Data relationships
• Regulatory readiness
Data Security
Customize
• Sanitization rules repository
• Setup stable and reliable environment
• Functional regression testing
• Build cross reference table for reusability
Maintenance
• Security level enhancement
• Critical production support
Governance

Figure 2.

The data sanitization offering consist of four parts, viz. (i) Data security consulting; (ii) Sanitization solution customization and implementation; (iii) Maintenance and support and (iv) Governance.

The Productized offering methodology is depicted in Figure 2. The data sanitization solution was to be implemented across various organizations and hence the architecture of the data sanitization solution needed to be very scalable. The underlying architecture of the Data Sanitization solution was organized into the four major components. It is important to note here that the methodology of putting in scalable service solution architecture is the most important step towards Productisation of service.

VII. Selling The Product

The product was launched in US market, since the organisation already had experience with more than 350 of the fortune 1000 companies and gathered deep domain knowledge in some of the important business areas. A team consisting of onsite and offshore individuals created the product offer and another sales team was made for USA in specific.

The sales team was involved into the productized service development to provide the market input from USA market. Series of trainings were conducted to articulate the features of this product and how it is designed to help customer address data security issues.

VIII. Summary And Conclusion

Repeatability is one of the keys to achieving scalable financial performance in the professional services firm because it improves service delivery consistency and thereby improves:

- Client satisfaction
- Project economics (better estimating, reduced learning curves among service delivery professionals, and so on)
- Practice economics (better predictability across the portfolio of projects)

By productizing and associating tangible features to an offering, the professional services firm can help ensure more consistent service delivery to realize these benefits. Predetermined templates for work products and deliverables, standardized methodologies, and fixed pricing and staffing models are all examples of standardized product attributes that can be assigned to service offerings. [Radford, 15]

Value addition made by professional services firms are solving client problems and no client problem is ever exactly the same. As a result, the actual service delivered for every client is customized to some extent.

However, significant benefits are gained by the professional services firm by associating product features with its services offerings—performance improvement can span sales, marketing, service delivery, and economics. These benefits are beneficial to professional services firms that find themselves positioned toward the latter half of the professional services lifecycle.

Productisation of services is accomplished largely by associating tangible features with intangible service offerings. These tangible features may take the form of personnel, collateral, methodologies, pricing, facilities, or other attributes. By associating tangible features with intangible services, the professional services firm can build client confidence during the sales cycle as well as during the service delivery phase.

The findings of this study shows that the company rolled out data sanitization as a productized service offering, packaged around the sanitization repository and their knowledge of customization for the local USA market need. This has gained a ground in the USA market especially in their current customers. The rollout of data sanitization in the USA geography in financial services industry has been successful so far and the company plans to replicate the same in the other geographies as well. In addition to the concrete changes in the services, the Productisation project provided results that have a more general meaning and probably long-lasting effects on the orientation of the companies. First of all, the attitudes towards Productisation changed, and secondly, the organization developed the skills of product development and also

References

[1] KUCZMARSHI T. D. & JOHNSTON Z. T. (2005). “Service Development”. Kenneth B. Kahn (Eds.), The PDMA handbook of new product development (pp.92-107). New Jersey: John Wiley & Sons, Inc

[2] SHETH, J. N., & SHARMA, A. (2008). “The impact of the product to service shift in industrial markets and the evolution of the sales organization”. Industrial Marketing Management. doi:10.1016/j.indmarman.2007.07.010

[3] VALMINEN, K., & TOIVONEN, M. (2007). “Improving competitiveness and performance through service Productisation? A case study of small KIBS companies participating in a Productisation project”. Service Engineering and Management Summer School (SEM 2007). Helsinki University of Technology. September 10

[4] DE BRENTANI (U.), 1991, “Success factors in Developing New Business Services”, European Journal of Marketing, Vol. 25, No. 2, 33-59

[5] GRÖNROOS (C.), 1990, “Service Management and Marketing”, Lexington, MA and Toronto, Lexington Books

[6] SUNDBO (J.), GALLOUJ (F.), 2000, “Innovation as a Loosely Coupled System in Services”, in Metcalfe (J.S.), Miles (I.), Innovation Systems in the Service Economy - Measurement and Case Study Analysis, Boston, Dordrecht and London, Kluwer Academic Publishers

[7] JAAKKOLA (E.), ORAVA (M.), VARJONEN (V.), 2007, “Competitiveness through Productisation. Guide to the companies”, Helsinki, Tekes (In Finnish)

[8] EDVARDSSON (B.), 1997, “Quality in new service development: Key concepts and a frame of reference”, International Journal of Production Economics, Vol. 52, 31-46

[9] VAATTOVAARA (M.), 1999, “Transforming services into products in a systems engineering companies”, Espoo, Helsinki University of Technology, Industrial Management and Work and Organizational Psychology, Report No 9

[10] GALLOUJ (F.), 2002, “Knowledge-Intensive Business Services: Processing Knowledge and Producing Innovation”, in Gadrey (J.), Gallouj (F.), Productivity, Innovation and Knowledge in Services. New Economic and Sosio-Economic Approaches, Cheltenhamn, Edward Elgar

[11] SIPILÄ (J.), 1999, “The Productisation of expert services”, Porvoo, WSOY (In Finnish)

[12] S. Zhong, Z. Yang, and R. Wright, “Privacy-Enhancing k-Anonymization of Customer Data,” Proceedings of the 24rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 139–147 (2005).

[13] CARR JASON, DOUG MARKIEWICZ, “Guidelines for Data Sanitization and disposal”, http://www.cmu.edu/iso/governance/guidelines/data-sanitization.html (October 2007)

[14] SHRINGARPURE VAIBHAV, “Compliance to Regulations: A Data warehouse approach”, http://www.information-management.com/infodirect/20050318/1023201-1.html (March 2005)

[15] RADFORD JOEL. (2004). “Service Productisation”, Microsoft Corporation and Epicor Software Corporation.

[16] GARTNER ANALYSIS report, http://www.gartner.com/it/products/research (2008)

.

Subpixel Accuracy Analysis of Phase Correlation Shift Measurement Methods Applied to Satellite Imagery

S.A. Mohamed, A.K. Helmi
Division of Data Reception Analysis and Receiving Station Affairs, National Authority for Remote Sensing and Space Sciences, Cairo, Egypt

M.A. Fkirin, S.M. Badwai
Department of Industrial Electronic Engineering and Control, Faculty of Electronic Engineering, MenoufiaUniversity, Menoufia, Egypt

***Abstract*—the key point of super resolution process is the accurate measuring of sub-pixel shift. Any tiny error in measuring such shift leads to an incorrect image focusing. In this paper, methodology of measuring sub-pixel shift using Phase correlation (PC) are evaluated using different window functions, then modified version of (PC) method using high pass filter (HPF) is introduced . Comprehensive analysis and assessment of (PC) methods shows that different natural features yield different shift measurements. It is concluded that there is no universal window function for measuring shift; it mainly depends on the features in the satellite images. Even the question of which window is optimal of particular feature is generally remains open. This paper presents the design of a method for obtaining high accuracy sub pixel shift phase correlation using (HPF).The proposed method makes the change in the different locations that lack of edges easy.**

Keywords-phase correlation (PC); high pass filter (HPF); window function; sub-pixel shift.

I. Introduction

Sub pixel image shift estimation is a fundamental task for high performance image processing techniques such as image fusion and super resolution, which have been extensively used for applications in remote sensing, medical imaging, surveillance and computer vision. In [1], frequency based shift calculated methods using phase correlation (PC) have been widely used because of its accuracy and low complexity for shift motion due to translation, rotation or scale changes between images. The PC method for images alignment relies mainly on the shift property of the Fourier transform to estimate the translation between two images [2]. It is extended to estimate rotation and scale changes by using log-polar coordinate changes [3]. Originally limited to discover only integer pixel translations, the algorithm can be naturally extended to provide sub pixel accuracy [4], but at a higher computational cost. On the other hand, newer approaches have also allowed obtaining sub pixel accuracy with much less complexity [5], and some of its latest variations have reported enhanced accuracy performances [6, 7, 8].

Using phase correlation with window; the window function plays a crucial role in determining the phase (wave front) because it significantly influences phase error. Window functions are used in harmonic analysis to reduce the undesirable effects related to the spectral leakage. They impact on many attributes of a harmonic processor which include detestability, resolution, dynamic range, confidence and ease of implementation [9]. Several standard windows are also used to optimize the requirements of a particular application in signal processing. Window functions have been successfully used in various areas of signal processing and communications such as, spectrum estimation, speech processing, digital filter design, and in other related fields such as, beam forming. A complete review of many window functions and their properties was presented by Harris [10].

All window functions are designed to reduce the side lobes of the spectral output of Fast Fourier transform (FFT) routines. Whilst applying the window function reduces the side lobe leakage, it causes the main lobe to broaden thus, reducing the resolution. This is a trade-off that has to be made, one should choose the window function, which best suits the application. Some windows which have small side-lobe level and quick fall off rate have been used to reduce the measurement error of dielectric loss factor tan δ caused by non-synchronized sampling and non-integral period truncation. The limitations of Fast Fourier transform (FFT) application in measurements are due to spectral leakage and the picket-fence effect. Spectral leakage is typically reduced by selection of the proper window [10]. The picket-fence effect errors are compressed by interpolated FFT. The FFT interpolation formula for the rectangular window was introduced in [11]. It was then extended for a Hanning window in [12].It is also known that windows with a narrow main lobe have better noise immunity [13]. In this paper we have used some different window functions for evaluation and applied them on 3 different locations in satellite images.

II. Subpixel Registration By Phase Correlation Method

To obtain the phase correlation function. Let the image I_2 be a shifted version of the image I_1 by$(x0, y0)$, then

$$I_2(x, y) = I_1(x - x0, y - y0) \qquad (1)$$

After taking the Fourier Transform (FT) of both images, we have the following relationship due to the shift property of the FT

$$\mathrm{I}_2(u,v) = \mathrm{I}_1(u,v) \;\; e^{-j(ux0+vy0)} \qquad (2)$$

Therefore, a shift in the spatial domain will produce a phase difference in the frequency domain. The normalized cross power spectrums finally denudes

$$\frac{I_2(u,v)I_1(u,v)}{|I_2(u,v)I_1(u,v)|} = e^{-j(ux0+vyo)} \qquad (3)$$

The (PC) function is finally obtained by taking the Inverse Fourier Transform (IFT) of the cross-power spectrum, which gives a $\delta\,(\mathrm{x0},\mathrm{y0})$ as result: a Dirac function centered on the position $(\mathrm{x0},\mathrm{y0})$.

None the less, appointed out in [5], when dealing with discrete images and using the Fast Fourier Transform (FFT) to generate the PC, the Dirac is turned into a Dirichlet kernel, whose maximum peak is found at the closest integer displacement, so finding the PC peak is equivalent to finding the translation at a pixel resolution. In order to obtain sub pixel resolution and keep using the same technique offending the peak position of the PC function, interpolation by zero padding the cross-power spectrum is suggested in [4], but accuracy is limited by the interpolation factor used which is also limited by the size of the IFFT that can be computed. Lately, this approach has been improved with a more efficient implementation proposed in [14]. Using a different approach, in[5] an extension of the original PC method is presented where using not only the PC information of the main peak, but also its surrounding pixels, leads to an estimation of the amount of sub pixel displacement as well. Let $\mathrm{C}(0,0)$ be the main peak, and $\mathrm{C}(1,0)$ and $\mathrm{C}(0,1)$ be the neighbors with the largest value in both horizontal and vertical direction respectively. The sub pixel displacement is then calculated as:

$$\Delta x = \frac{c(1{,}0)*xp+c(0{,}0)*x}{c(1{,}0)+c(0{,}0)} - 1 \quad \Delta y = \frac{c(0{,}1)*yp+c(0{,}0)*y}{c(0{,}1)+c(0{,}0)} - 1 \qquad (4)$$

Where x_p positive X direction and y_p positive Y direction.

III. WINDOWS FUNCTION & HIGH PASS FILTER BASED ON PHASE CORRELATION

A. *Window Function*

In signal processing, a window function (also known as an apodization function or tapering function [15]) is a mathematical function that is zero-valued outside of some chosen interval. For instance, a function that is constant inside the interval and zero elsewhere is called a rectangular window, which describes the shape of its graphical representation. When another function or a signal (data) is multiplied by a window function, the product is also zero-valued outside the interval: all that is left is the part where they overlap; the "view through the window". Applications of window functions include spectral analysis, filter design, and beam forming. The following table describes the window functions [15].

TABLE I. WINDOWS FUNCTIONS.

Window	**Functions**
Triangle	$w[n] = \frac{2}{N+1}.\left(\frac{N+1}{2} - \left\lvert n - \frac{N-1}{2}\right\rvert\right)$
Hanning	$w[n] = 0.5(1 - cos\left[\frac{2\pi n}{N}\right])$
Hamming	$w[n] = 0.5 + (1-0.5)cos\left[\frac{2\pi n}{N}\right]$
Kaiser	$w[n] = \frac{I0\left(3\pi\sqrt{1-(\frac{2n}{N-1})^2}\right)}{I0(3\pi)}$
Chebwin	$w[n] = \frac{cos\left(N*\cos^{-1}\left[\alpha*\cos\left(\frac{\pi m}{n}\right)\right]\right)}{\cosh[N*\cosh^{-1}(\alpha)]}$
Blackman	$w[n] = 0.42 + 0.5cos\left[\frac{2\pi n}{N}\right] + 0.08cos\left[\frac{4\pi n}{N}\right]$
Boxcar	$w[n] = RECT\left[\frac{n}{0.97N}\right]$
Bartlett	$w[n] = \frac{2}{N-1}.\left(\frac{N-1}{2} - \left\lvert n - \frac{N-1}{2}\right\rvert\right)$

B. *High Pass Filter*

A high pass filter (HPF) could be analog or digital filter. Analog filter should be electronic circuits to filter image before recording, on the other hand digital filter will deal with the image after being recorded, so here we are using digital filter. Filter could be Finite impulse response (FIR) or Infinite impulse response (IIR) [16, 17]. The main difference is that FIR would have a linear phase but IIR will give non linear phase. In our method we need to make calculations on phase, so we have more concern in FIR to get better result and fast processing. For instance, a HPF will be used to reduce low frequency parts details in the image as desert is considered as the low frequency component in the data of the image.

IV. MATERIALS AND METHOD

A. *Data Sets*

The French satellite SPOT5 was launched in 2002. It has a resolution of 5 m for the panchromatic band (HRG instrument) with an unchanged swath of 60 km [18]. Extracting the 2.5 m from 5m images requires a special technique. We will use tow scenes; their characteristics are shown in Table 2.

TABLE II. IMAGE CHARACTERISTICS.

Images	**Image_1**	**Images_2**
K	113	113
J	289	289
Description	Panchromatic	Panchromatic
Wavelength(μm)	0.48-0.71	0.48-0.71
Spectral mode	A	B
Processing level	1A	1A
Acquisition date	16/06/2008	16/06/2008

B. Operation Steps

Steps for phase correlation using window function & high pass filters are as follows:

- Read image & apply windows and high pass filter.
- Get 2D FFT for both images.
- Get phase correlation surface, defines as $C_{t,t+1}(k,l) = F^{-1}(\frac{F_t^* F_{t+1}}{|F_t^* F_{t+1}|})$ where Ft and Ft+1 are respectively the two-dimensional Fourier transforms of F_t and F_{t+1}, F^{-1} denotes the inverse Fourier transform and* denotes complex conjugate.
- Get maximum real point in 2D surface.
- Get position of this point x, y.

  $$(K_m , l_m) = argmaxRe\{C_{t,t+1}(k,l)\}$$

  Where (k_m ,l_m) co-ordinates of the maximum value of array $C_{t,t+1}(k,l)$
- Calculate sub pixel difference Δx and Δy from eq. 4.

The algorithm starts from the already shifted image. The roughly coregistered images (to the same reference system spot5) are correlated using a small sliding window. We apply a different window function on both images to reduce noise edge effects. We calculate the Fast Fourier transform FFT of both images. Then, Calculate the cross-power spectrum by taking the complex conjugate of the result. By multiplying the Fourier transforms together, element wise, we obtained phase correlation surface by applying the inverse Fourier transform. Finally we determined the location of the maximum peak from real value of surface(x, y). The entire processing chain is summarized in Figure.1.

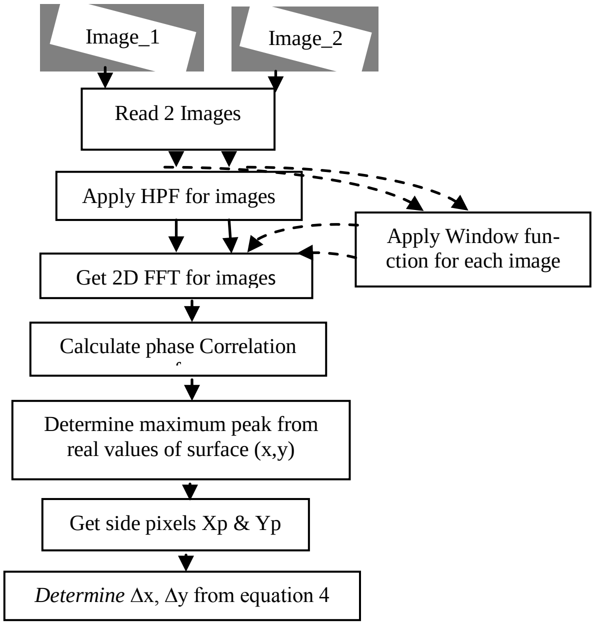


Image_1
Image_2
Read 2 Images
Apply HPF for images
Apply Window function for each image
Get 2D FFT for images
Calculate phase Correlation
Determine maximum peak from real values of surface (x,y)
Get side pixels Xp & Yp
Determine Δx, Δy from equation 4

Figure .1 Algorithm workflow diagram of the proposed phase correlation based window function method and high pass filter

V. Data analysis

We applied the previous algorithm In order to simulate the sub pixel displacements using a couple of panchromatic images (2000×2000 pixels). The application was conducted using different windows, on shifted images as shown in Figure.2. These two images have"-0.5, 0.5 pixel" shift between them. There are many phenomena in the images that will affect the shift results as shown in Table.3.

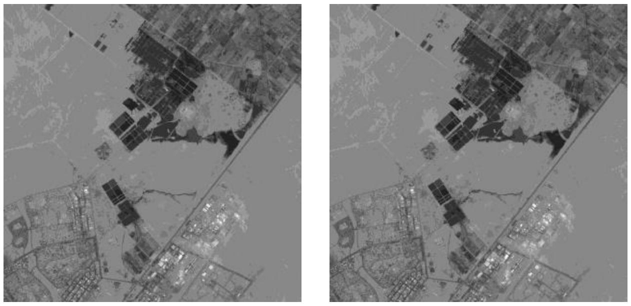


Figure.2. Two sample images shifted by"-0.5, 0.5" pixel.

TABLE III. Output widows shift results of the source image.

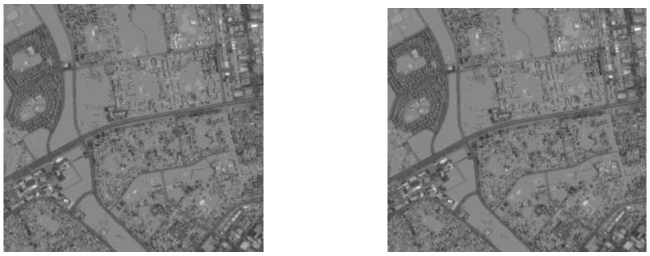
windows	Source image	
	Δx	Δy
No Window	-0.7450	0.7303
Triangle	-0.7784	0.7617
Hanning	-0.7874	0.7621
Hamming	-0.7819	0.7603
Kaiser	-0.8028	0.7771
Chebwin	-0.8056	0.7798
Blackman	-0.7986	0.7731
Boxcar	-0.7450	0.7303
Bartlett	-0.7784	0.7617

VI. Evaluation of different area locations

This section presents the evaluation of some windows functions on three different locations namely; (urban area, vegetation area and desert area).

A. Urban area

This area was chosen from the above data of Table. 2. The choice of this region was done because of the large number of urban areas, which have a lot of edges that makes the process faster in measuring the displacement between the two images. Figure.3 shows the selected area (500×500) pixels, with different buildings and Figure.4 depicts the urban pixel values with pixel location.



(a) (b)

Figure.3.Two urban areas presenting roads, buildings and sharp edges.

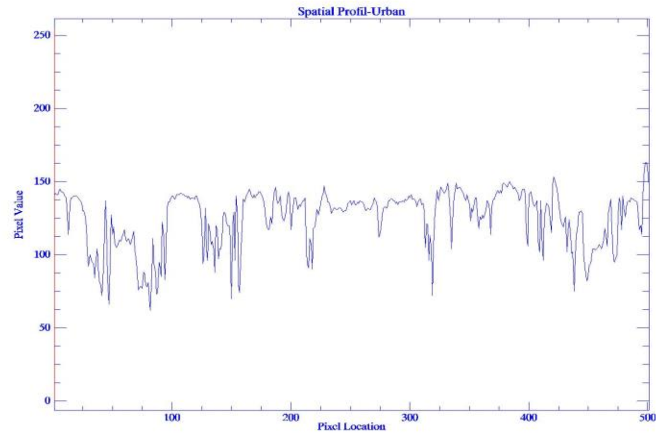
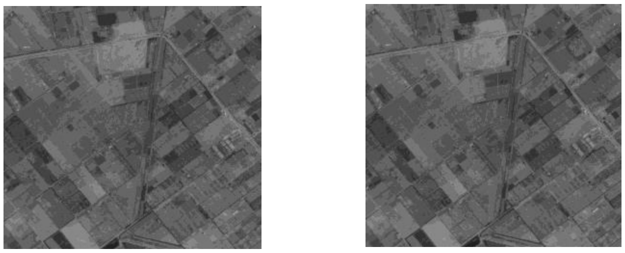


Figure.4. Result of urban pixel values with pixel location.

B. Vegetation area

This area was chosen from the above data of Table. 2. The Choice of this region was done because of the large number of Agricultural fields, in which the edges appear between the different parts in the fields, but not in abundance, such as buildings. Also the image size of these areas is (500×500) pixel.



(a) (b)

Figure.5. Two vegetation areas presenting fewer edges than urban.

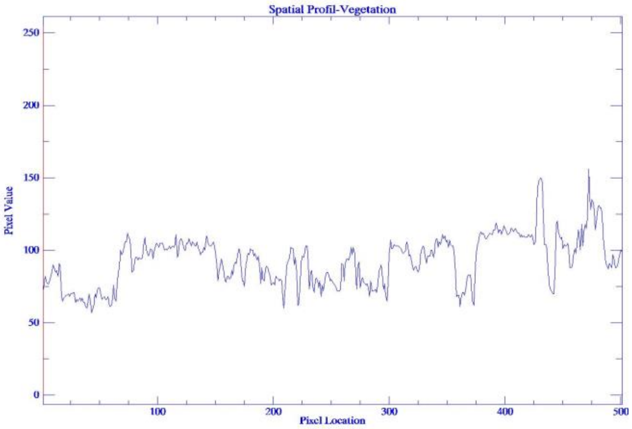
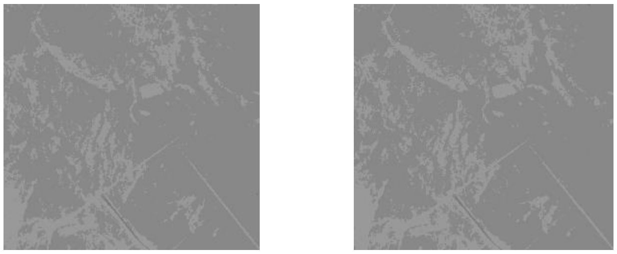


Figure.6. Result of vegetation pixel values with pixel location.

Figure.5 shows the vegetations area and Figure.6 depicts the vegetation pixel values with pixel location.

C. Desert area

This area was also chosen from the above data of Table 2. This desert region does not have edges that help successful completion of the measurements process. Also the image size of this area is (500x500) pixel. Figure.7 shows the smoothed desert area and Figure.8 depicts the desert pixel values with pixel location.



(a) (b)

Figure.7. Two desert areas presenting no edges.

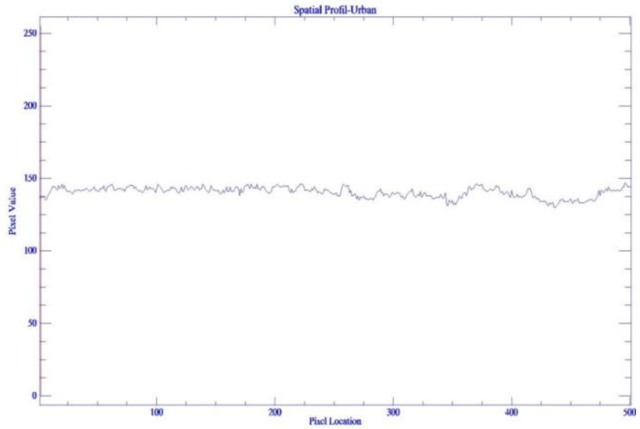


Figure.8. Result of desert pixel values with pixel location.

VII. RESULTS AND DISCUSSION

A. Urban area

We applied the algorithm discussed in section 4 on the subset images of the buildings and roads. It has been found after using 8 different windows that; all windows are giving variable and different results from each other, but closer to the (actual shift) as shown in Table. 4, thanks to the large amount of edges produced by buildings.

B. Vegetation area

Also in this part the algorithm was applied to the fields of agriculture using the 8 different windows yielding poor result, as shown in Table.5. These results are due to the lack of edges that are on the expense of displacement calculation accuracy between the two images. However, as shown in the table without the use of windows, it gives a bad result, as well as using the Box windows.

C. Desert area

In case of the desert, we used the 8 different windows on areas where there are no edges which do not lead to satisfactory results and complete the measurement process successfully. Figure.8 shows pixel values and there is no change that results from the edges. Therefore, results shown in Table.6 are very bad for all the windows.

D. High pass filter (HPF)

Finally we applied HPF (FIR filter) to overcome the problem of the desert, which has very bad results. Table.7 shows the results of using the filter on the different locations. It is clear that it gives a very comparable result.

TABLE IV. TRESULTS OF THE DIFFERENT WINDOWS IN URBAN AREA, ΔX AND ΔY ARE VALUES REPRESENTING THE SUB PIXEL SHIFT CALCULATED FROM THE DISCUSSED METHOD.

Windows	Urban	
	Δx	Δy
No window	-0.5255	0.5057
Triangle	-0.5408	0.5501
Hanning	-0.5488	0.5573
Hamming	-0.5444	0.5507
Kaiser	-0.5589	0.5661
Chebwin	-0.5608	0.5674
Blackman	-0.5569	0.5643
Boxcar	-0.5255	0.5057
Bartlett	-0.5409	0.5503

TABLE V. RESULT OF THE DIFFERENT WINDOWS IN THE VEGETATION AREA.

windows	Vegetation	
	Δx	Δy
No window	-0.2587	0.4673
Triangle	-0.6861	0.7974
Hanning	-0.6910	0.7954
Hamming	-0.6878	0.7979
Kaiser	-0.6879	0.7908
Chebwin	-0.6864	0.7881
Blackman	-0.6903	0.7920
Boxcar	-0.2587	0.4673
Bartlett	-0.6863	0.7971

TABLE VI. RESULT OF THE DIFFERENT WINDOWS IN THE DESERT AREA.

windows	Desert	
	Δx	Δy
No window	0.0518	0.4511
Triangle	-1.0692	1.0961
Hanning	-1.0804	1.0996
Hamming	-1.0729	1.0988
Kaiser	-1.0890	1.0973
Chebwin	-1.0885	1.0984
Blackman	-1.0880	1.0988
Boxcar	0.0518	0.4511
Bartlett	-1.0695	1.0959

TABLE VII. RESULTS OF THE DIFFERENT LOCATION USING HPF.

Δ	Source image	Urban area	Vegetation area	Desert area
Δx	-0.5014	-0.5001	-0.5093	-0.4053
Δy	0.5140	0.5014	0.5137	0.5013

VIII. CONCLUSION

In this paper we present a algorithm for obtaining high-accuracy sub-pixel shit estimation using phase correlation. It appeared that having a large number of pixels gives a good representation of Fourier and using windows is not giving a better correlation solution in low frequency component (Desert). Results of applying the window functions did not succeed in enhancing the calculation of shift in images, but the (HPF) did make enhancement on source image by removing low frequency components hence helping us to calculate peak phase shift.

REFERENCES

[1] B. Zitova and J. Flusser, “Image registration methods: a survey,” Image and Vision Computing, vol.21, no.11, pp.977–1000, October 2003.

[2] C. D. Kuglin and D. C. Hines, “The phase correlation image alignment method,” in Proc. Int. Conference on Cybernetics and Society, pp. 163–165, 1975.

[3] B. Reddy and B. Chatterji, “An fft-based technique for translation, rotation, and scale-invariant image registration,” IEEE Trans. on Image Processing, vol. 5, no. 8, pp. 1266–1271, 1996.

[4] B. Marcel, M. Briot, and R. Murrieta, “Calcul de translation et rotation par la transformation de fourier,” Traitement du Signal, vol. 14, no. 2, pp. 135–149, 1997.

[5] H. Foroosh, J. Zerubia, and M. Berthod, “Extension of phase correlation to subpixel registration,” IEEE Trans. on Image Processing, vol. 11, no. 3, pp. 188–200, 2002.

[6] K. Takita, T. Aoki, Y. Sasaki, T. Higuchi, and K. Kobayashi, “High-accuracy subpixel image registration based on phase-only correlation,” IEICE Trans. Fund., vol. E86-A, no. 8, pp. 1925–1934, 2003.

[7] V. Argyriou and T. Vlachos, “Performance study of gradient correlation for sub-pixel motion estimation in the frequency domain,” IEE Proc. - Vision, Image, and Signal Processing, vol. 152, no. 1, pp. 107–114, 2005.

[8] V. Argyriou and T. Vlachos, “On the estimation of subpixel motion using phase correlation,” Journal of Electronic Imaging, vol. 16, no. 033018, 2007.

[9] Y.H. Ha, J.A. Pearce," A new window and comparison to standard windows," IEEE Trans. Acoust., Speech Signal Process., 37, vol no.2, pp. 298–301,1989.

[10] F. J. Harris, “On the use of windows for harmonic analysis with the discrete Fourier transform,” Proc. IEEE, vol. 66, no. 1, pp. 51–83, Jan. 1978.

[11] V. H. Jain, W. L. Collins, and D. C. Davis, “High accuracy analog measurement via interpolated FFT,” IEEE Trans. Instrum. Meas., vol. IM-28, no. 2, pp. 113–122, Jun. 1979.

[12] T. Grandke, “Interpolation algorithms for discrete Fourier transform of weighted signals,” IEEE Trans. Instrum. Meas., vol. IM-32, no. 2, pp. 350–355, Jun. 1983.

[13] K. duda " DFT Interpolation Algorithm for Kaiser–Bessel and Dolph–Chebyshev Windows" IEEE Transaction on Instrumentation and Measurement, vol. 60, no. 3, March 2011.

[14] M. Guizar-Sicairos, S. T. Thurman, and J. R. Fienup, “Efficient subpixel image registration algorithms,” Optics Letters, vol. 33, no. 2, pp. 156–158, 2008.

[15] L. M. Surhon, M. T. Timpledon, S. F. Marseken,"window Function"ISBN 613030014X, 9786130300142, VDM Verlag, 2010, p.124.

[16] A. E. Cetin, O.N. Gerek, Y. Yardimci, "Equiripple FIR filter design by the FFT algorithm," IEEE Signal Processing Magazine, pp. 60-64, March 1997.

[17] Ranjit Singh, Sandeep K. Arya," Genetic Algorithm for the Design of Optimal IIR Digital Filters," Journal of Signal and Information Processing, pp. 286-292, August 2012.

[18] C. Fratter, M. Moulin, H. Ruiz, P. Charvet, D. Zobler, “The SPOT5 mission”, 52nd International Astronautical Congress, Toulouse, France, 1–5 Oct 2001.

Performance Comparison of Gender and Age Group Recognition for Human-Robot Interaction

Myung-Won Lee
Dept. of Control and Instrumentation Engineering,
Chosun University, 375 Seosuk-dong
Gwangju, Korea

Keun-Chang Kwak*
Dept. of Control, Instrumentation, and Robot Engineering,
Chosun University, 375 Seosuk-dong
Gwangju, Korea

***Abstract*—In this paper, we focus on performance comparison of gender and age group recognition to perform robot's application services for Human-Robot Interaction (HRI). HRI is a core technology that can naturally interact between human and robot. Among various HRI components, we concentrate audio-based techniques such as gender and age group recognition from multichannel microphones and sound board equipped with robots. For comparative purposes, we perform the performance comparison of Mel-Frequency Cepstral Coefficients (MFCC) and Linear Prediction Coding Coefficients (LPCC) in the feature extraction step, Support Vector Machine (SVM) and C4.5 Decision Tree (DT) in the classification step. Finally, we deal with the usefulness of gender and age group recognition for human-robot interaction in home service robot environments.**

Keywords-gender recognition; age group recognition; human-robot interaction.

I. INTRODUCTION

Conventional industrial robots perform jobs and simple tasks by following pre-programmed instructions for humans in factories. Meanwhile, the main goal of the intelligent service robot is to adapt to the necessities of life as accessibility to human life increases. While industrial robots have been widely used in many manufacturing industries, intelligent service robots are still in elementary standard. Although the intelligent robots have been brought to public attention, the development of intelligent service robots remains as a matter to be researched further.

Recently, there has been a renewal of interest in Human-Robot Interaction (HRI) for intelligent service robots [1-2]. This is different from HCI (Human-Computer Interaction) in that robots have an autonomous movement, a bidirectional feature of interaction, and diversity of control level. Among various HRI components, we especially focus on audio-based HRI. Audio-based HRI technology includes speech recognition, speaker recognition [3][4], sound source localization [5], sound source separation, speech emotional recognition, speech enhancement, gender and age group recognition. Among various audio-based HRI components, we focus on gender and age group recognition. The robot platform used in this paper is WEVER, which is a network-based intelligent home service robot equipped with multi-channel sound board and three low-cost condenser microphones. Finally, we perform the performance comparison in the step of feature extraction (MFCC, LPCC) and classification (SVM, C4.5) for gender and age group classification.

The material of this paper is organized into following fashion. In section 2, we describe and discuss about well-known feature extraction methods Mel-Frequency Cepstral Coefficients (MFCC) and Linear Prediction Coding Coefficients (LPCC). In section 3, we deal with Support Vector Machine (SVM) and C4.5 Decision Tree (DT) for classification. Here we elaborate on gender and age group recognition in home service robots equipped with multiple microphones and multi-channel sound board. In section 4, we perform the experimental setup and performance comparison. Finally the conclusions and comments are given in section 5.

II. FEATURE EXTRACTION METHODS

The speech signals are obtained from the first channel of sound board at a distance of 1 meter in quite office environments. The speech signal is sampled with 16kHz, and each sample is encoded with 16bits. There are 20 sentences in a long speech signal for gender classification data, only one sentence in a speech signal for age classification data. Speech signals are assumed to be time invariant within a time period of 10 to 30 ms. Short-time processing methods are adopted for speech signal processing. A window sequence is used to cut the speech signal into segments, and short-time processing is periodically repeated for the duration of the waveform. The key problem in speech processing is to locate accurately beginning and ending of a speech. Endpoint detection (EPD) enables computation reduction and better recognition performance. We detect beginning and ending points of speech intervals using short-time energy and short-time zero crossings

The short time energy is as follows

$$E_n = \sum_{m=-\infty}^{\infty} \left[x(m) w(n-m) \right]^2 \tag{1}$$

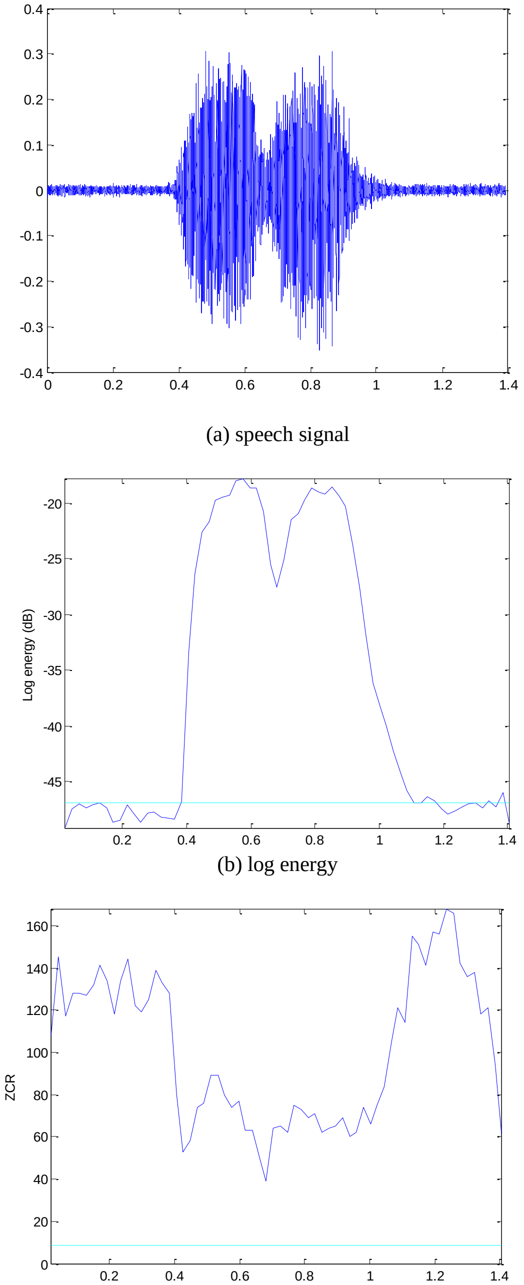
The short time zero crossing rate is as follows

$$Z_n = \sum_{m=-\infty}^{\infty} \left| sgn\left[x(m) \right] - sgn\left[x(m-1) \right] \right| w(n-m) \tag{2}$$

Figure 1 shows the EPD obtained from log energy and zero crossing rate. Rectangular window gives equal weights to all samples in the window. Hamming window gives most weight to middle sample. Rectangular and Hamming windows can be expressed as follows, respectively.

$$w(n)=1,\quad 0\le n\le N \ \ and \ \ 0 \ \ otherwise \tag{3}$$

$$w(n)=\begin{cases}0.54-0.46\cos(2\pi n/(N-1)) & 0\le n\le N-1\\ 0 & otherwise\end{cases} \tag{4}$$



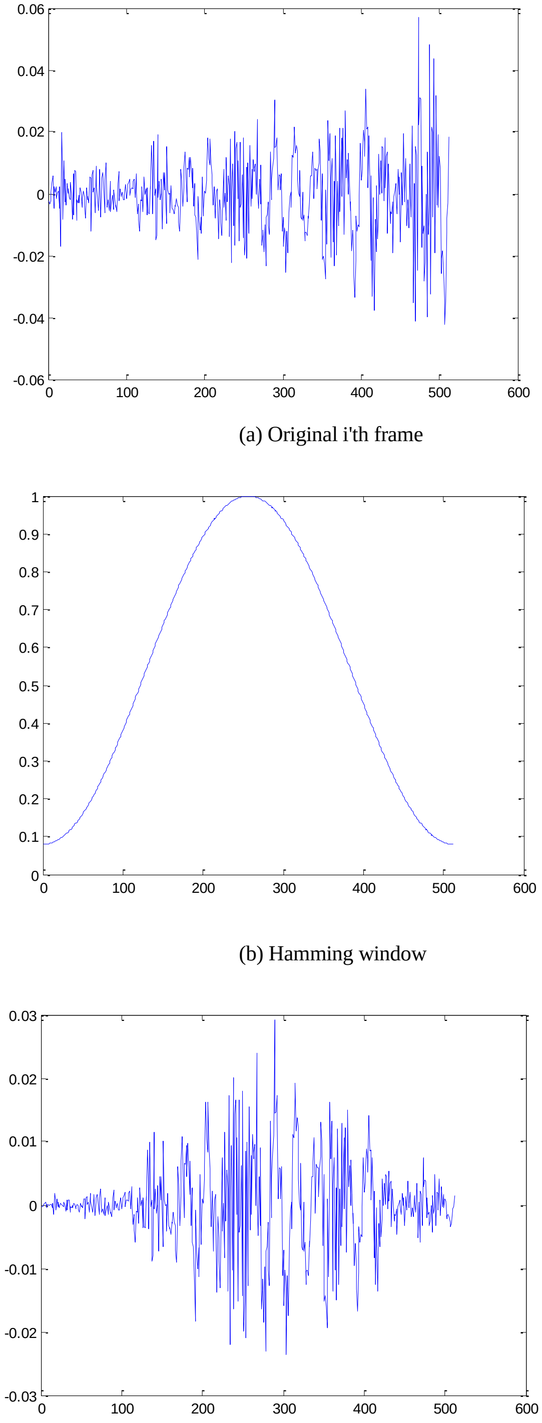
(a) speech signal

(b) log energy

(c) zero crossing rate

Figure 1. Endpoint detection

The window used is hamming window, and the window length is 512 samples with 30% overlap. Figure 2 shows original i'th frame and the transformed i'th frame obtained by hamming window.



(a) Original i'th frame

(b) Hamming window

(c) i'th frame obtained by hamming window

Figure 2. Signal obtained by Hamming window

After detecting endpoints of the speech interval, silence intervals from the speech signal are removed. We cut the speech signal into segments, each segment is 512 sample points in each frame for the signal with 16kHz sample rate. The size of overlapped frame is 171 samples.

The number of the filter bank is 20. The dimension of MFCC is 12. Feature extraction is based on each frame of the speech signals. After detecting signal, the feature extraction step is performed by six stages to obtain MFCC. These stages consist of pre-emphasis, frame blocking, hamming window, FFT (Fast Fourier Transform), triangular bandpass filter, and cosine transform [6]. For simplicity, we use 11 MFCC parameters except for the first order. The construction procedure of MFCC is shown in Figure 3.

The mel scale filter bank is a series of triangular bandpass filters hat have been designed to simulate the bandpass filtering believed to occur in the auditory system. This corresponds to a series of bandpass filters with constant bandwidth and spacing on a mel frequency scale as shown in Figure 4.

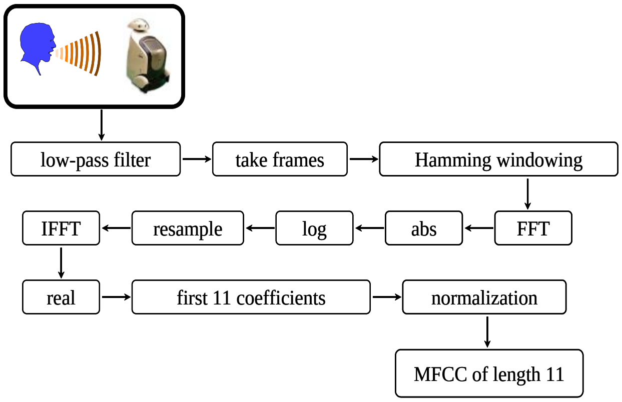


Figure 3. Procedure of MFCC.

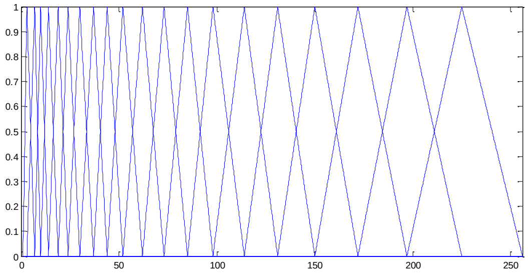


Figure 4. Mel scale filter bank

On the other hand, LPCC (Linear Prediction Coding Coefficients) method provides accurate estimates of speech parameters. The current speech sample can be closely approximated as a linear combination of past samples.

$$x(n) \approx a_1 x(n-1) + a_2 x(n-2) + \cdots + a_p x(n-p) \tag{5}$$

Coefficients are determined by minimizing the sum of squared differences between the actual speech samples and the linearly predicted ones.

III. Classification Methods

We firstly consider the use of the support vector machine (SVM) as a nonlinear classifier. This is legitimated by the fact that SVMs come with high generalization capabilities. Among various SVM models, we use the one known as LIBSVM. LIBSVM is composed of C-support vector classification (C-SVM) and ν -support vector classification (ν -SVM). Here we employ the C-SVM in the form proposed by Vapnik [7] for the implementation of multi-class classification. Furthermore we consider polynomial kernel functions frequently used in conjunction with classification tasks

Polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$ (6)

where γ and r are kernel parameters.

On the other hand, C4.5 is a method used to generate a decision tree developed by Quinlan [8]. C4.5 is an extension of ID3 algorithm. The decision tree generated by C4.5 can be used for classification. For this reason, C4.5 is often referred to as a statistical classifier. C4.5 builds decision tree from a set of training data in the same way as ID3 using the concept of information entropy. The training data is a set S=s_1, s_2, … of already classified samples. Each sample s_i = x_1, x_2, … is a vector where x_1, x_2, … represents attributes or features of the sample. The training data is augmented with a vector C = c_1, c_2, … where c_1, c_2, … represent the class to which each sample belongs. At each node of the tree, C4.5 choose one attribute of the data that most effectively splits its set of sample into subsets enriched in one class or the other. Its criterion is the normalized information gain that results from choosing an attribute for splitting the data.

The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recourses on the smaller sublists. The algorithm has a few base cases. All the samples in the list belongs to the same class. When this happens, it simply creates a leaf node of the decision tree saying to choose that class. None of the features provide any information gain. In this case, C4.5 creates a decision node higher up the tree using the expected value of the class. Instance of previously-unseen class encountered. Again, C4.5 creates a decision node higher up the tree using the expected value.

IV. Experimental Results

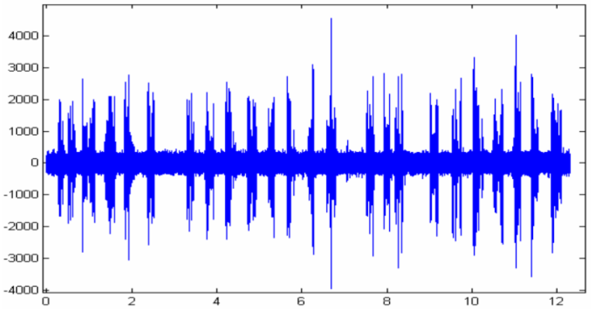
In this section, we describe our comprehensive set of experiments and draw conclusions regarding the classification performance in comparison with well-known methods frequently used in conjunction with the feature extraction and classification. We used hamming window 512 samples to multiply the speech signal to enable short-time speech signal processing. We extract MFCC and LPCC features based on each frame to produce feature data for classification.

The classification data includes 14 long speech signals for gender classification from 7 female and 7 male, 500 speech signals (200 signals from children, 300 signals from adults) for age group classification, respectively. We divide 2/3 of examples for training into the rest for testing. The training and testing data for gender classification is 6960 and 3482, respectively.

The training and testing data for age group classification is 12925 and 6464, respectively. Figure 5 shows 20 sentences in a long speech signal for gender classification data. Figure 6 shows u-robot test bed environments including three rooms and a living room. Figure 7 shows multi-channel sound board. These microphones are low-price condenser. Furthermore, multi-channel sound board was developed for sound localization and speech/speaker recognition in Electronics Communications Research Institutes (ETRI). Figure 8 visualizes MFCC obtained from one sentence.

Table 1 lists the result of performance comparison for gender classification. As listed in Table 1, the experimental results revealed that MFCC-SVM showed good performance (93.16%) in comparison to other presented approaches for testing data set. Table 2 lists the result of performance comparison for age group classification. The experimental results obtained that MFCC-SVM (91.39%) outperformed other methods in like manner.

As a result, the SVM and DT classifiers obtained better performance with MFCC features than LPCC features. Auditory model has been introduced in the MFCC feature, other auditory model embedded features can be extracted for future classification. Other features as pitch period, formants, short-time average energy etc. can be extracted to combine with MFCC or LPCC features for classification.



Speech signals with 20 sentences

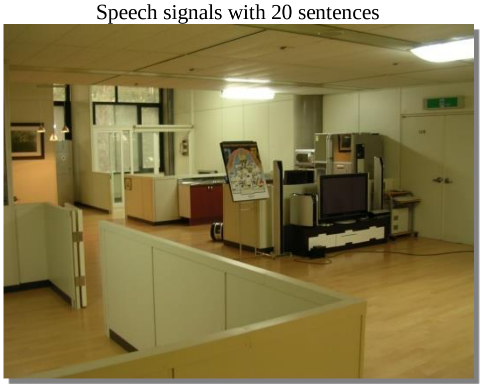


Figure 5. Robot test bed environment

Figure 6. Multi-channel sound board(MIM board)

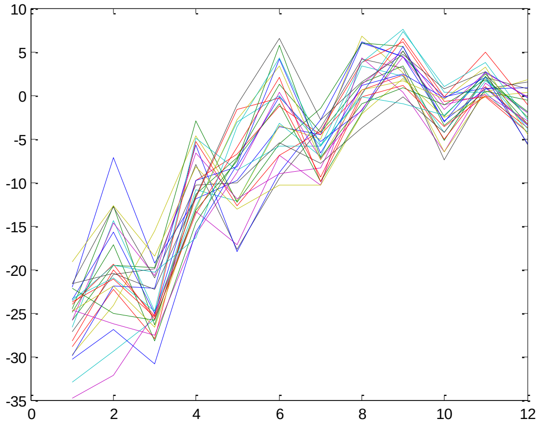


Figure 7. MFCC obtained by one sentence.

TABLE I. PERFORMANCE COMPARISON (GENDER CLASSIFICATION)

	Training data	Testing data
MFCC-SVM	95.89	93.16
MFCC-DT	94.33	91.45
LPCC-SVM	93.28	86.60
LPCC-DT	93.10	83.02

TABLE II. PERFORMANCE COMPARISON (AGE GROUP CLASSIFICATION)

	Testing data
MFCC-SVM	91.39
MFCC-DT	88.37
LPCC-SVM	84.69
LPCC-DT	82.72

V. CONCLUSIONS

We have performed the comparative analysis for gender and age group classification of audio-based HRI components. These components are compared with MFCC-SVM, MFCC-DT, LPCC-SVM, and LPCC-DT. The experimental results revealed that the aggregate of the MFCC-SVM showed better performance in comparison with other methods. We have shown the usefulness and effectiveness of the presented technique through the performance obtained from the constructed databases.

In the future studies, we shall continuously develop other techniques such as sound source separation and fusion of information obtained from multi-microphones for humanlike robot auditory system. Also, we can apply to customized service application based on the integrated robot audition system including gender and age group recognition, speaker and speech recognition, and sound source localization and separation.

The presented technique can be applied to service robots such as home service robots, edutainment robots, and u-health robots as well as various application areas.

REFERENCES

[1] K. C. Kwak, S. S. Kim, "Sound Source Localization With the Aid of Excitation Source Information in Home Robot Environments," IEEE Trans. on Consumer Electronics, Vol. 54, No. 2, 2008, pp. 852-856.

[2] K. C. Kwak, "Face Recognition with the Use of Tensor Representation in Home Robot Environments", IEICE Electronics Express, Vol. 6, No. 4, 2009, pp. 187-192.

[3] M. Ji, S. Kim, and H. Kim, "Text-independent speaker identification using soft channel selection in home robot environments," IEEE Consumer Electronics, vol. 54, no. 1, 2008, pp.140-144.

[4] K. S. R. Murty and B. Yegnanarayana, "Combining evidence from residual phase and MFCC features for speaker recognition," IEEE Signal Processing Letters, vol. 13, no. 1, 2006, pp. 52-55.

[5] V. C. Raykar, B. Yegnanarayana, S. R. M. Prasanna, and R. Duraiswami, "Speaker localization using excitation source information in speech," IEEE Speech and Audio Processing, vol. 13, no. 5, 2005, pp.751-761.

[6] D. A. Reynolds, R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models", IEEE Trans. on Speech and Audio Processing, vol. 3, no. 1, 1995, pp. 72-83.

[7] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

[8] J. R. Quinlan, C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993

AUTHORS PROFILE

Myung-Won Lee received the B.Sc. and M.Sc. from Chosun University, Gwangju, Korea, in 2010 and 2012, respectively. He is currently pursuing a candidate for the Ph.D. His research interests include human–robot interaction, computational intelligence, and pattern recognition.

Keun-Chang Kwak received the B.Sc., M.Sc., and Ph.D. degrees from Chungbuk National University, Cheongju, Korea, in 1996, 1998, and 2002, respectively. During 2003–2005, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. From 2005 to 2007, he was a Senior Researcher with the Human–Robot Interaction Team, Intelligent Robot Division, Electronics and Telecommunications Research Institute, Daejeon, Korea. He is currently the Assistant Professor with the Department of Control, Instrumentation, and Robot Engineering, Chosun University, Gwangju, Korea. His research interests include human–robot interaction, computational intelligence, biometrics, and pattern recognition. Dr. Kwak is a member of IEEE, IEICE, KFIS, KRS, ICROS, KIPS, and IEEK.

Spatial Cloud Detection and Retrieval System for Satellite Images

Noureldin Laban[1], Ayman Nasr[1]
[1]Department of Image Processing and its Applications,
Data Reception and Analysis Division, National
Authority for Remote Sensing and Space Sciences ,
P.O.Box 1564, Alf Maskan, Cairo, Egypt

Motaz ElSaban[2] Hoda Onsi[2]
[2]Department of Information Technology,
Faculty of Computers and Information, Cairo University,
P.O.Box 12613 , Orman, Giza, Egypt

***Abstract*—In last the decade we witnessed a large increase in data generated by earth observing satellites. Hence, intelligent processing of the huge amount of data received by hundreds of earth receiving stations, with specific satellite image oriented approaches, presents itself as a pressing need. One of the most important steps in earlier stages of satellite image processing is cloud detection. Satellite images having a large percentage of cloud cannot be used in further analysis. While there are many approaches that deal with different semantic meaning, there are rarely approaches that deal specifically with cloud detection and retrieval. In this paper we introduce a novel approach that spatially detect and retrieve clouds in satellite images using their unique properties .Our approach is developed as spatial cloud detection and retrieval system (SCDRS) that introduce a complete framework for specific semantic retrieval system. It uses a Query by polygon (QBP) paradigm for the content of interest instead of using the more conventional rectangular query by image approach. First, we extract features from the satellite images using multiple tile sizes using spatial and textural properties of cloud regions. Second, we retrieve our tiles using a parametric statistical approach within a multilevel refinement process. Our approach has been experimentally validated against the conventional ones yielding enhanced precision and recall rates in the same time it gives more precise detection of cloud coverage regions.**

Keywords-Satellite images; Content based image retrieval; Query by polygon; Retrieval refinement; cloud detection; geographic information system.

I. INTRODUCTION

Satellite images have become a common component of our daily life either on the Internet, in car driving and even in our hand-held mobile handsets. There is huge image content appearing every second through multiple competing satellite systems [1]. Manual interaction with this large volume of data is becoming more and more inappropriate, which creates an urgent need for automatic treatment to store, organize and retrieve this content [2].

Traditional textual meta-data such as geographic coverage, time of acquisition, sensor parameters, manual annotation, etc., are now insufficient to retrieve images of interest when we target a specific visual concept such as desert, rock, crops, clouds or others [3]. In many fields, we need specific contents from the satellite images as specific crops, geology structures or climate changes.

Manual annotation needs to annotate every region by human where users enter descriptive word after image download from satellite. However it is a labor intensive and tedious process [4]. Therefore we need to retrieve images that contain our intended contents automatically. The content based image retrieval (CBIR) approach challenge is how to fill the gap between the low level features that describe the scenes and our human understandable semantic concepts. This gap of understanding is called the semantic gap [5] [6]. In addition, these semantic concepts themselves may be defined differently, e.g. each one of us interprets what he sees from his point of view.

The most commonly used features include those reflecting color, texture, shape, and salient points in an image. For instance, in a color layout approach, an image is divided into a small number of sub-images and the average color components (e.g. red, green, and blue intensities) are computed for every sub-image [7]. Texture features are intended to capture the granularity and repetitive patterns of surfaces within an image.

The traditional satellite cloud image search method was based on the file name and the sensor parameters of every image. The disadvantages of this method are that it cannot describe the image contents such as cloud shape [8] and also leads to the inconvenience in retrieving images [9].

We have done statistics for Spot4 satellite observation on the Middle East from NARSS archive to determine the percent of clouds on these scenes in the period starts from January 2006 to December 2009. There were about 170000 scenes covering the receiving station area. Normally for each scene; an expert has to decide manually the percentage of cloud coverage.

The different percentages of clouds coverage during each year are shown in figure 1 and table I.

TABLE I : AVERAGE CLOUD COVERAGE THOUGH 2006 TO 2009 ON MIDDLE EAST

COVERAGE	2006	2007	2008	2009
0% (A)	0.44	0.39	0.43	0.40
1%-10% (B)	0.08	0.09	0.07	0.06
11%-25% (C)	0.09	0.08	0.07	0.05
26%-75% (D)	0.09	0.08	0.07	0.06
76%-100% (E)	0.30	0.36	0.36	0.43

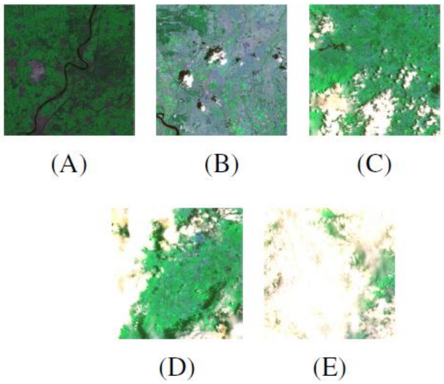


Fig. 1 Clouds coverage percentages

II. Review Of Related Work

During the last decade many approaches have been proposed to retrieve satellite images using their content in general. Specifically less effort has been devoted to cloud despite its importance during satellite image processing or meteorological management and observation. F. Acqua and P. Gamba presented a tool for shape similarity evaluation for query-by shape searching into meteorological image archives based on the point diffusion technique [8]. R. Holowczak et al., reported a system that can automatically determine whether a region of interest is visible in the image, free from cloud, and can incorporate this into the meta-data for individual images to enhance searching capability [10]. T. Nauss et al., have proposed an algorithm based on the analytical solutions of the radiative transfer equations valid for optically thick weakly absorbing cloud layers [11]. D. Fu and L. Xu have used 2D-Gabor wavelet in satellite image classification [12]. D. Upreti has used Gray level Co-occurrence Matrix GLCM and histogram quantization technique to retrieve cloud patterns to discover Tropical Cyclone [13].

The previous approaches were concerned with cloud retrieval. Some observations were found as follow:

- Most of the previous work was directed to meteorological observation images with very low resolution.
- It doesn’t care with cloud removal preprocessing operation which is still done manually.
- It doesn’t handle spatial distribution of cloud within the scene.

Through our new proposed approach, we covered these missed points of research. It will be very useful to detect and retrieve these clouds and consequently as further process, remove them and replace the cloudy sub-images with other clear ones.

III. System Overview

Our system is composed of two main stages as shown in figure 2.

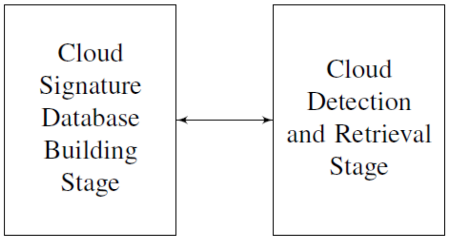

Cloud Signature Database Building Stage
Cloud Detection and Retrieval Stage

Fig. 2 System Overview

First stage is cloud signature database building stage which is responsible for building up the features vectors for different clouds patterns. Second stage is cloud detection and retrieval stage in each satellite scene, which determines where the clouds in this scene and their percentage are. We have used two strategies in our system [1]. First one is query by polygon strategy where we build our signature database using cloud polygons instead of rectangular shapes. Second one is multiple size tiling strategy where we break down our scene into different sizes followed by features extraction to obtain features vectors. According to these strategies, the two stages have passed through different sub-processes starting by tiling then features extraction to from features vectors. This is done for each level of retrieval.

IV. System Stages

A. Cloud Signature Database Building stage

There are many forms that clouds appear with in satellites images as shown in figure 3.

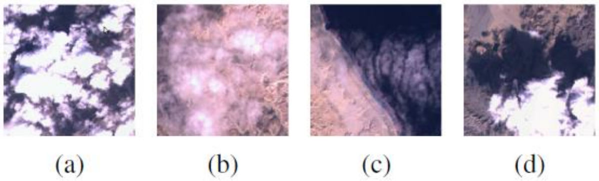


Fig. 3 Some Clouds types

These forms differ depending on altitude and density of clouds [14]. These forms start with low dense water vapor to high dense clouds with different altitude. Beside clouds there are also their shadows which should be taken into account during retrieval. The first stage of the cloud retrieval process is to determine cloud signature as shown in figure 4.

This is done using query by polygon approach where we first determine different type of clouds, then we draw geo-reference polygons that contain these clouds. These different types of clouds are used to form signature databases according to the type of tiling size used. Using our proposed feature extraction algorithm we compute features vectors of cloud polygon tiles

B. Cloud Detection and Retrieval Stage

After building our cloud signature database, we have to build the features vectors for each scene as shown in figure 5.

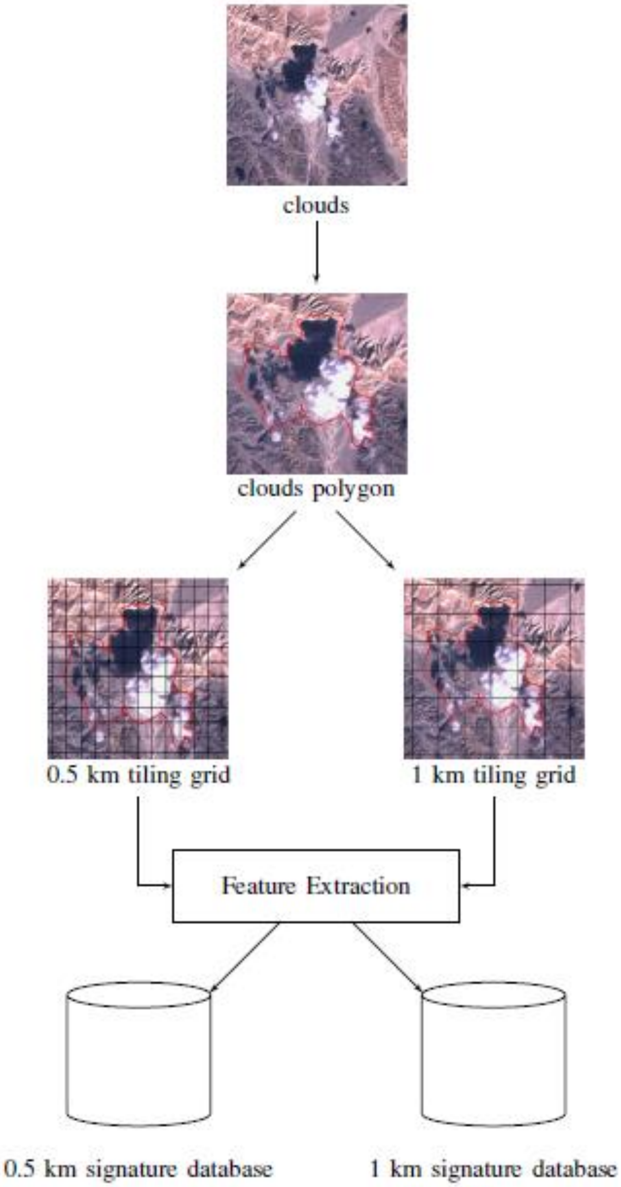


Fig. 4 Building cloud signature databases

Our approach based on breaking down the whole image into small sections of sub-images called tiles. The number of resulted tiles is determined by their sizes.

According to the two stages hierarchy used in [1] for the retrieval process, we have rebuilt the system. Instead of starting with features databases and get query features for each semantic, we have reverse the order which begins with building cloud signature database then the input scene is treated as query image. The two stages hierarchy, candidate selection stage and refinement stage, are used. In candidate selection stage, we define the primary candidate's area for clouds. In refinement stage we refine the first stage areas using its neighborhoods with smaller tile size.

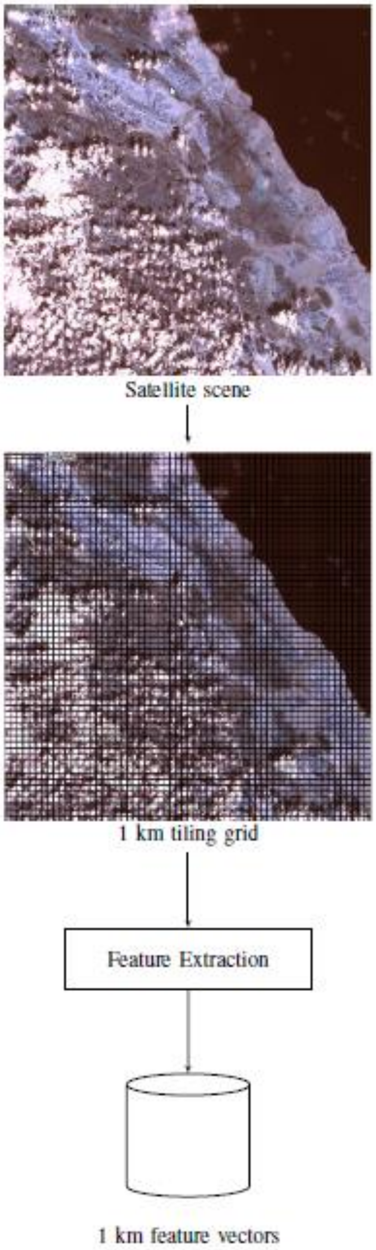


Fig. 5 Building satellite feature vectors

V. Main System Processes

The main system processes; features extraction, retrieval and evaluation have some key points to be included into the two levels hierarchy to enhance cloud detection and retrieval system.

A. Features extraction Process

We have depended on various domains to get the tile signature either for the cloud example dataset or for input satellite image. These domains extract the spectral and textural characteristics of images.

To build our feature vector database $\mathrm{DB}_{\mathrm{features}}$, we start by determining the components of our feature vector $\mathrm{V_i}$ for each tile i and its length. For each multispectral tile image $\mathrm{T_i}$ with number of bands, we form the feature vector $\mathrm{V_{ib}}$ of each band b depending on different spectral and textual characteristics of the image. We used the mean μ and standard deviation σ statistics of feature domain for each band. The features we used are histogram H, Daubechies wavelets transform coefficients DWT,Discrete cosine transform coefficients DCT and Discrete Fourier transform Coefficient DFT [15]. Using these domains, we build various feature vectors $\mathrm{V_H}$, $\mathrm{V}_{\mathrm{DCT}}$ and $\mathrm{V_{DFT}}$. For each multispectral tile with n number of bands, we build domain feature vector $\mathrm{V_d}$ for each domain d as in equation 1.

$$V_d = [V_{d1}, V_{d2}, \dots\dots\dots\dots.., V_{dn}] \qquad (1)$$

We then use these domain feature vectors to form domain feature database DB_d for m number of tiles as in equation 2.

$$DB_d = [V_{d1}, V_{d2}, \dots\dots\dots\dots.., V_{Dm}] \qquad (2)$$

Using all feature vectors for all tiles; we formulate our cloud signature database or input scene feature vectors using all domains as in equation 3.

$$DB_{feature} = [DB_1, DB_2, \dots\dots\dots.., DB_d] \quad (3)$$

B. Retrieval Process

The retrieval process, as shown in figure 6, has two sub stages as mentioned in [1], the candidates selection stage and the refinement stage.

In the candidates selection stage, we use 1 km tile size features to get the most appropriate matching tiles similar to cloud. In the refinement stage, we use the 0.5 km tile size features of the first stage results and their neighborhoods to get our final results.

We have used a retrieval engine that based on statistical parametric paradigm using normal distribution [16] rather than the traditional nearest neighbor approach. The statistical parametric paradigm aimed to determine the parameters of the statistical distribution that the data follows as mean μ and standard deviation σ . We define the the training dataset D_{training} that represent cloud example tiles set D_{cloud} and non cloud example tiles set $D_{\mathrm{non\ cloud}}$ as in equation 4.

$$D_{training} = [D_{cloud}, D_{non\ cloud}] \qquad (4)$$

This is done for every tile size. Therefore, our global signature data D_{global} is formed from all sizes used in our system as in equation 5.

$$D_{global} = [D_{size\ 1}, D_{size\ 2}, \dots\dots, D_{last\ size}] \quad (5)$$

After we have built our statistical model using D_{training}, SCDRS is now ready to receive the satellite images as an input.

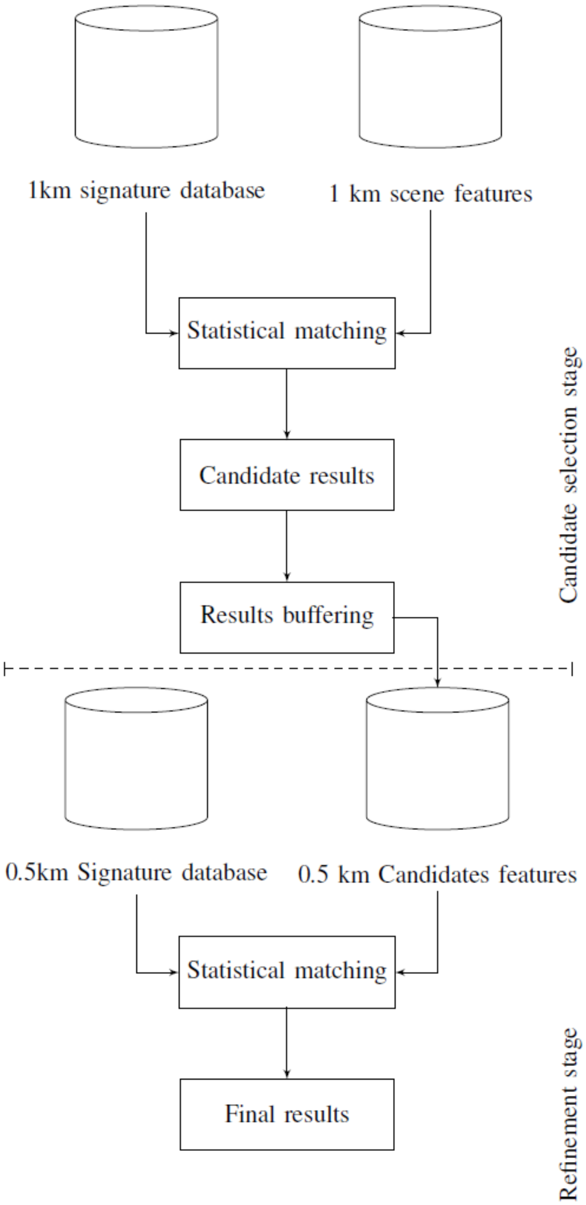


Fig. 6 Two Stages Retrieval Process

C. System Evaluation Process

Our evaluation process is carried out in terms of recall and precision (equations 6, 7 respectively) using relevant areas in the database.

$$recall = \frac{\text{correctly retrieved cloud area}}{\text{actual cloud area}} \qquad (6)$$

$$\text{precision} = \frac{\text{correctly retrieved cloud area}}{\text{retrieved cloud area}} \quad (7)$$

We use the map coordinates (i.e. Latitude and Longitude) instead of using file coordinates (pixels). As the map coordinates is universal and continuous where the file coordinates is file specific. The global coordinate system is independent from the pixel size whatever the scanning satellite or stored file. So the percent of cloud area in the input scene is as shown in equation 8

$$\text{retrieved cloud percent} = \frac{\text{retrieved cloud area}}{\text{whole scene area}} \quad (8)$$

where the actual cloud percent retrieved is calculated as shown in equation 9

$$\text{actual cloud percent} = \frac{\text{correctly retrieved cloud area}}{\text{whole scene area}} \quad (9)$$

VI. Experimental Results

On our experiments we have used Spot4 satellite scenes with different cloud cover percents which cover about $10800\ km^2$. Each scene covers $60\ km \times 60\ km$ of earth surface in Egypt with pixel size of $20m$. We used also Landsat archive images database with different cloud coverage percentages. There scenes cover about $22400\ km^2$ with $30\ m$ pixel size.

Each scene has been divided into sub images of $1\ km \times 1\ km$ and $0: 5\ km \times 0: 5\ km$. The experiment scenes have formed more than $100{,}000$ sub-images which are pre-classified clouds images. We have used samples of different clouds types to form our cloud signature database which is composed of 110 sub images acting as clouds examples.

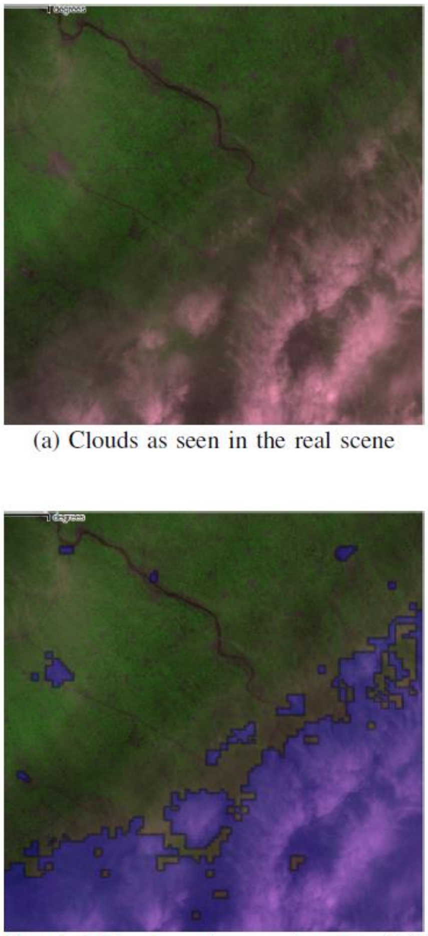
VII. Results Analysis

For our semantic concept which is cloud; first we have used two categories of polygons shapes, one used for building cloud signature database and the other is tied with each input scene used for evaluation. An example result of our system is depicted in figure 7.

The results of each input scene could be evaluated by two ways. First, the input test polygon for cloud; which determines exactly the positions of clouds in this scene and the area of clouds compared to the whole scene area. Second, the expert's estimation used in ground station which estimate the range of cloud cover as explained in table I.

As shown in table II, results of the two successive stages of the system are presented. It shows how the different types of features domains affect the results.

To determine cloud percentage coverage, we have calculated the total area of output results cloud tiles with respect to the whole scene area which is $3600\ km^2$ as in equation 8. We have put into consideration that the most important parameter is precision as we should guarantee that the output results have to be more accurate and decrease the non clouds tiles resulted. So, when we select the cloud examples, it should be purely determined.



(a) Clouds as seen in the real scene

(b) Detected Clouds regions retrieved by SCDRS

Fig. 7 SCDRS result example

TABLE II : DIFFERENT RECALL AND PRECISION FOR TWO STAGE HIERARCHY

	first stage		second stage	
	Precision	Recall	Precision	Recall
Histogram	76	88	72	91
Wavelets	74	91	73	92
DCT	74	90	72	92
FFT	72	87	70	93

Table III shows the recall and precision results using the different feature domains. The accuracy of different features is very comparable. The results explain that the key point here is the processing time, which is recorded to histogram features as it is the least complex than the others. As the tile becomes more smaller the spectral characteristics become more sufficient than textural characteristics to distinguish between tiles.

TABLE III : DIFFERENT RECALL(R) AND PRECISION (P) FOR DIFFERENT TYPES OF FEATURES USING 0.5 KM TILE SIZE AND PROCESSING TIME (PT)

	Histogram		Wavelets		Discrete cosine		Fourier	
	P	R	P	R	P	R	P	R
A	88	94	89	100	84	100	89	100
B	66	90	68	90	67	86	68	90
C	86	75	75	75	80	75	79	69
D	97	82	97	76	97	79	97	79
E	100	97	100	97	100	97	100	100
PT(MIN)	22		45		28		24	

VIII. Conclusions

In this paper, a new approach was developed to detect the percentage of clouds and retrieve their positions within the satellite images using two stages; Cloud Signature Database Building stage and Cloud Detection and Retrieval Stage. The two stages used multilevel framework hierarchy of candidates selection and candidates refinement processes. This is done using spatial and textural features and parametric statistical approach for retrieval process. The capability of the developed system was tested using a dedicated satellite images and assessed in terms of cloud percentage coverage with the traditional precision and recall measurements. Results show that the developed system enhanced the precision and recall and in the same time it gives a closer assessment for cloud coverage to the real area calculations. They also show that the spectral features have higher accuracy than textural features. We propose as future work to represent a system for detecting different types of clouds using more robust retrieval algorithms which integrated with GIS systems.

References

[1] N. Laban, M. ElSaban, A. Nasr, and H. Onsi, “System refinement for content based satellite image retrieval,” The Egyptian Journal of Remote Sensing and Space Sciences, vol. 15, June 2012.

[2] M. Martins, L. Frutuoso Guimaraes, and L. Maria Garcia Fonseca, “Texture feature neural classifier for remote sensing image retrieval systems,” in Computer Graphics and Image Processing, 2002. Proceedings. XV Brazilian Symposium on, 2002.

[3] C.-R. Shyu, M. Klaric, G. Scott, A. Barb, C. Davis, and K. Palaniappan, “Geoiris: Geospatial information retrieval and indexing system mdash;content mining, semantics modeling, and complex queries,” Geoscience and Remote Sensing, IEEE Transactions on, vol. 45, no. 4, pp. 839 –852, April 2007.

[4] H. H. Wang, D. Mohamad, and N. A. Ismail, “Semantic gap in cbir: Automatic objects spatial relationships semantic extraction and representation,” International Journal Of Image Processing, vol. 4, pp. 192–286, July 2010.

[5] I. Gondra and D. R. Heisterkamp, “Content-based image retrieval with the normalized information distance,” Computer Vision and Image Understanding, vol. 111, no. 2, pp. 219 – 228, 2008.

[6] H. Min and Y. Shuangyuan, “Overview of content-based image retrieval with high-level semantics,” in 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), 2010, vol. 6, Aug. 2010, pp. 312 –316.

[7] R. Datta, D. Joshi, J. Li, and J. Z. Wang, “Image retrieval: Ideas, influences, and trends of the new age,” ACM Comput. Surv., vol. 40, no. 2, pp. 5:1–5:60, May 2008.

[8] F. Dell’Acqua and P. Gamba, “Query-by-shape in meteorological image archives using the point diffusion technique,” Geoscience and Remote Sensing, IEEE Transactions on, vol. 39, no. 9, pp. 1834 –1843, Sept. 2001

[9] W. ShangGuan, Y. Hao, Y. Tang, and Y. Zhu, “The research and application of content-based satellite cloud image retrieval,” in International Conference on Mechatronics and Automation. ICMA 2007., Aug. 2007, pp. 3864 –3869.

[10] R. Holowczak, F. Artigas, S. A. Chun, J.-S. Cho, and H. Stone, “An experimental study on content-based image classification for satellite image databases,” Geoscience and Remote Sensing, IEEE Transactions on, vol. 40, no. 6, pp. 1338 – 1347, June 2002.

[11] T. Nauss, A. Kokhanovsky, T. Nakajima, C. Reudenbach, and J. Bendix, “The intercomparison of selected cloud retrieval algorithms,” Atmospheric Research, vol. 78, no. 12, pp. 46 – 78, 2005.

[12] D. Fu and L. Xu, “Satellite cloud image texture feature extraction based on gabor wavelet,” in Image and Signal Processing (CISP), 2011 4th International Congress on, vol. 1, Oct. 2011, pp. 248 –251.

[13] D. Upreti, “Content-based satellite cloud image retrieval,” Master’s thesis, Faculty of Geo-Information Science and Earth Observation of the University of Twente , Enschede, The Netherlands, 2011.

[14] J. Oliver, Ed., Encyclopedia of World Climatology, ser. Encyclopedia of Earth Sciences Series. Springer, 2005.

[15] M. Petrou and C. Petrou, Image Processing: The Fundamentals. Wiley, April 2010.

[16] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: a review,” ACM Computer Survey, vol. 31, no. 3, pp. 264–323, Sept. 1999.

Equalization: Analysis of MIMO Systems in Frequency Selective Channel

Amit Grover[1*]
Assistant Professor, Department of ECE
Shaheed Bhagat Singh State Technical Campus,
Moga Road (NH-95), Ferozepur -152004, India.

Abstract—**Due to the increased demand of wireless communication systems because of the features of the system which provides a wide coverage, high throughput and reliable services, the MIMO systems communication has come into existence. Features provided by these systems ensure the improved system coverage and increased data transmission rate by considering multiple numbers of transmitter and receiver antennas. In this paper, we are considering the equalization; a filtering approach that minimizes the error between actual output and desired output by continuous updating its filter coefficients for Rayleigh Frequency selective fading channel. We concluded that MMSE and ZF give the worst performance in Rayleigh frequency selective channel as compare to Rayleigh Flat fading Channel [35] due to a constant BER for large SNR's. We have also observed that the successive interference methods provide better performance as compare to others, but their complexity is high. ML provides the better performance in comparison to others and BER doesn't remain constant for a large SNR in this case. Simulation results shows that ML equalizer with BPSK gives better performance as compare to QPSK. Finally we concluded that Sphere decoder provides the best performance.**

Keywords-Quadrature Amplitude Modulation (QAM); Quadrature Phase Shift Key (QPSK); Binary Phase Shift Key (BPSK); Minimum mean-squared error (MMSE); Maximum likelihood (ML); Bit error rate (BER); Inter-symbol interference (ISI); Successive-interference-cancellation (SIC); Sphere Decoder (SD); Zero Forcing (ZF).

I. Introduction

Because of the features of MIMO systems, it became an important part of modern wireless communication [5]. Communication in wireless channels is impaired predominantly by multipath fading. Multipath is the arrival of the transmitted signal at the receiver through differing angles and/or differing time delays and/or differing frequency [4]. MIMO offers significant increases in data throughput and link range without additional bandwidth or transmit power. It achieves this by higher spectral efficiency and link reliability and or diversity. The information bits to be transmitted are encoded and interleaved [9]. The Symbol mapper mapped the interleaved codeword's to data symbols and then these data symbols are applied as input to the Space Time-encoder which are again mapped to transmit antennas by space-time pre-coding block and are received at the antenna array by passing through the channel. The receiver performs the reverse operation to that of transmitter, followed by space time decoding [7].

II. Mimo System Model

By considering a MIMO system with a transmit array of $M_T$antennas and a receive array of $\mathrm{M_R}$ antennas [11], the MIMO channel [35] at a given time instant may be represented as a $M_R \times M_T$ matrix

$$\mathrm{H} = \begin{bmatrix} \mathrm{H_{1,1}} & \mathrm{H_{1,2}} & \cdots & \mathrm{H_{1,M_T}} \\ \mathrm{H_{2,1}} & \mathrm{H_{2,2}} & \dots & \mathrm{H_{2,M_T}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{H_{M_R,1}} & \mathrm{H_{M_R,2}} & \cdots & \mathrm{H_{M_R,M_T}} \end{bmatrix} \tag{1}$$

III. Mimo Channel

By assuming the above channel model the product of the bandwidth and the delay spread is very small and it results in channel realizations which varies with the time and is frequency dependent that is H (f) [35].

IV. Equalization Techniques

A. Zero forcing

Zero forcing is a linear equalization method that does not consider the effects of noise. In fact, the noise may be enhanced in the process of eliminating the interference. A zero-forcing equalizer uses an inverse filter to compensate for the channel response function [4].By assuming $\mathrm{M_T} = \mathrm{M_R}$ and H is a full rank square matrix, we can calculate the covariance matrix [35] of the effected noise as:

$$E[(n\mathrm{H}^{-1})^H . n\mathrm{H}^{-1}] = (H^{-1})^H . E[n^H . n]. H^{-1} = n(H.H^H)^{-1}$$

B. Minimum mean square error (MMSE)

The MMSE detector is the optimal detection that seeks to balance between cancelation of the interference and reduction of noise enhancement. We assume that the number of receive antennas is less than the number of transmit antennas $\mathrm{M} \geq \mathrm{N}$ MMSE at a high SNR [35] is given by

$$\mathrm{W_{MMSE}} = \left(\mathrm{H^*H} + \frac{1}{\mathrm{SNR}}\mathrm{I}\right)^{-1} \mathrm{H^*} \approx (\mathrm{H^H H})^{-1}\mathrm{H^H}$$

C. Successive Interference Cancelation.

The SIC schemes also reduces the noise amplification by the nulling vector. Therefore after the first cancelation the nulling vector for the second stream need only $\mathrm{M_r}$ -1 dimensions. The performance can also be enhanced by optimal ordering the SIC process, in which a nulling vector has been chosen that has the smallest norm to detect the corresponding data stream, but the system becomes more complicated.

D. Maximum Likelihood (ML)

Maximum likelihood detection [35] calculates the Euclidean distance between received signal vector and the product of all possible transmitted signal vectors with the given channel H, and finds the one with minimum distance.

The ML method achieves the optimal performance as the maximum a posterior detection when all the transmitted vectors are likely. However, its complexity increases exponentially as modulation order and/or the number of transmit antennas increases [6]. The ML receiver performs optimum vector decoding and is optimal in the sense of minimizing the error probability.

E. Sphere Decoder (SD)

By using sphere decoding, a limited number of codeword's are considered that are within a sphere centered at the received signal vector. By using this technique we find the ML solution vector that considers only a small set of vectors within a given sphere rather than all possible transmitted signal vectors [2]. By Considering the sphere with radius of R_{SD} as shown in equation (4)

$$(\bar{x} - \hat{\hat{x}})^T \bar{H}^T \bar{H} (\bar{x} - \hat{\hat{x}}) \le R_{SD}^2$$

SD method considers only the vectors inside a sphere defined by Equation (4). By considering the example of sphere which includes four candidate vectors, one of which is the ML solution vector. No vector outside the sphere can be the ML solution vector because their ML metric values are bigger than the ones inside the sphere [16]. If we were fortunate to choose the closest one among the four candidate vectors, we can reduce the radius in Equation (4) so that we may have a sphere within which a single vector remains. In other words, the ML solution vector is now constrained in this sphere with a reduced radius [35]. And we can express the Equation (4) as

$$\begin{aligned} = \left|r_{44}(\bar{x}_4 - \hat{\hat{x}}_4)\right|^2 + \left|r_{33}(\bar{x}_3 - \hat{\hat{x}}_3)\right|^2 + \left|r_{34}(\bar{x}_4 - \hat{\hat{x}}_4)\right|^2 \\ + \left|r_{22}(\bar{x}_2 - \hat{\hat{x}}_2)\right|^2 + \left|r_{23}(\bar{x}_3 - \hat{\hat{x}}_3)\right|^2 \\ + \left|r_{24}(\bar{x}_4 - \hat{\hat{x}}_4)\right|^2 + \left|r_{11}(\bar{x}_1 - \hat{\hat{x}}_1)\right|^2 \\ + \left|r_{12}(\bar{x}_2 - \hat{\hat{x}}_2)\right|^2 + \left|r_{13}(\bar{x}_3 - \hat{\hat{x}}_3)\right|^2 \\ + \left|r_{14}(\bar{x}_4 - \hat{\hat{x}}_4)\right|^2 \le R_{SD}^2 \end{aligned}$$

Using the Sphere in Equation (5), we consider a candidate value for $\hat{\hat{x}}_4$ in its own single dimension, which is arbitrarily chosen from the points in the sphere and this point, must be chosen in the following range:

$$\hat{\hat{x}}_4 - \frac{R_{SD}}{r_{44}} \le \bar{x}_4 \le \hat{\hat{x}}_4 + \frac{R_{SD}}{r_{44}}$$

If there exists no candidate point satisfying the inequalities, the radius needs to be increased. We assume that a candidate value was successfully chosen. Then we proceed to next step. We consider a candidate value for $\bar{x}_3$ now. If a candidate value for $\bar{x}_3$ does not exist, we go back to Step 1 and choose other candidate value of $\tilde{\tilde{x}}_4$ and search for $\bar{x}_3$ .In case that no candidate value $\bar{x}_3$ exists with all possible values $\tilde{\tilde{x}}_4$, we increase the radius of sphere R_{SD} , and repeat the step 1. Similarly a candidate value for $\bar{x}_1$ is chosen. After finding all candidate values, radius can be calculated by Equation (5).

Using this new radius Step 1 is repeated. If $[\tilde{\tilde{x}}_1 \ \tilde{\tilde{x}}_2 \ \tilde{\tilde{x}}_3 \ \tilde{\tilde{x}}_4]$ turns out to be a single point inside a sphere with that radius, it is declared as the ML solution vector [35] and searching procedure stops.

V. Simulation And Results

Using MATLAB, the different simulation results are shown in the different graphs, which provide the comparison of the BER for different modulation techniques using different equalizers like MMSE, ZF, ZF-SIC, MMSE-SIC and ML with Rayleigh frequency selective fading channel.

"Fig.4"and"Fig.5"shows the comparison of BER for different modulations with MMSE and ZF equalizers.

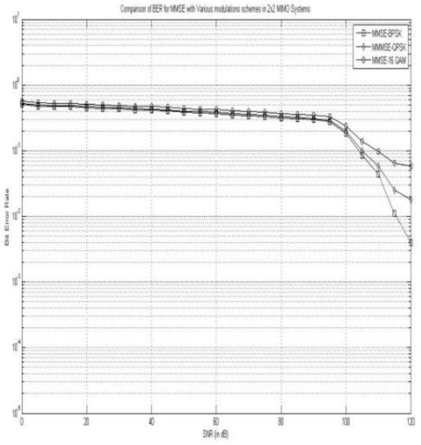


Figure 4.

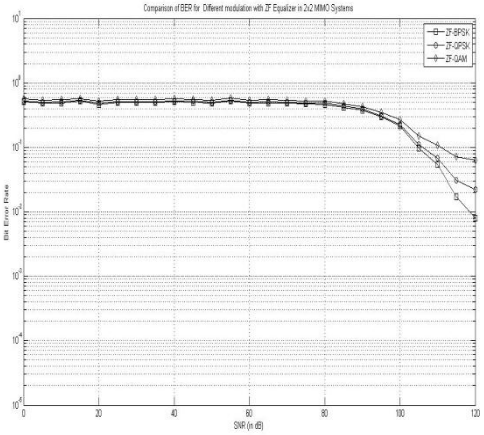


Figure 5.

The MMSE and ZF equalizer doesn't work well in Frequency selective fading channel. As it is clear from the "Fig.4" and "Fig.5" that BER remains constant for a large SNR and a deviation in BER is there at large SNR, and the performance of MMSE is little bit better in comparison to ZF. So the simulation of MMSE and ZF equalizer with BPSK by using MIMO 2 x2 Frequency selective fading channel, gives the better performance as compared to QPSK and 16QAM in both the cases.

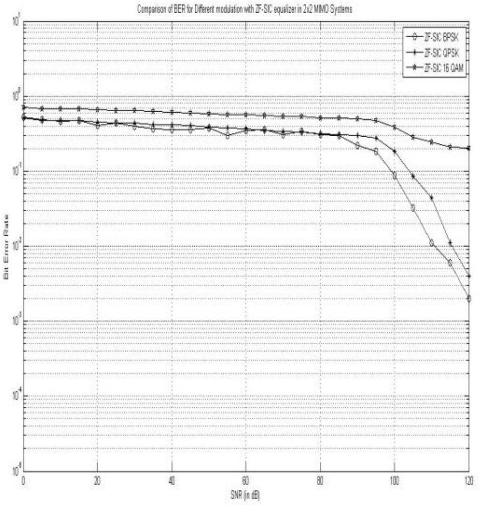


Figure 6.

"Fig.6"shows the comparison of BER for different modulations with ZF-SIC equalizer.

The ZF-SIC and MMSE-SIC equalizer doesn't work well in Frequency selective fading channel. As it is clear from the "Fig.6" and "Fig.7" BER remains constant for a large SNR and a deviation in BER is there at large SNR.

The performance of ZF-SIC is little bit better in comparison to ZF. So the simulation of ZF-SIC and MMSE-SIC equalizer with BPSK by using MIMO 2 x2 Frequency selective fading channel, gives the better performance as compared to QPSK and 16 QAM.

"Fig.7"shows the comparison of BER for different modulations with MMSE-SIC equalizer.

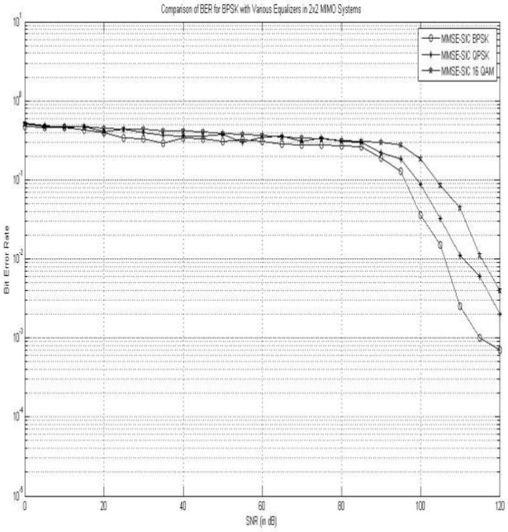


Figure 7.

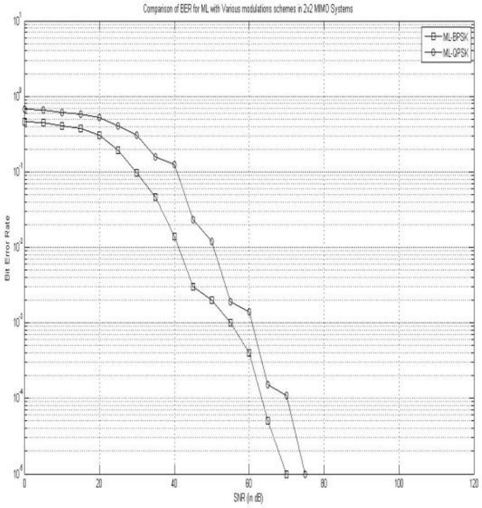


Figure 8.

"Fig.8"shows the comparison of BER for different modulations with ML equalizer.

The ML equalizer works well in Frequency selective fading channel in comparison to previously discussed equalizers. As it is clear from "Fig.8" BER doesn't remain constant for a large SNR as in another case. We can conclude that ML equalizer gives us good result in frequency fading. When we simulated ML equalizer with BPSK and QPSK by using MIMO 2 x2 Frequency selective fading channel, then as expected the performance of BPSK is better than QPSK.

"Fig.9"shows the comparison of all the equalizers with their best modulation scheme, and the simulation result for transmitting 2 bits/sec over two transmit and two receive antennas using BPSK. The results are decoded using the ZF, MMSE, ZF-SIC, MMSE-SIC, ML and Sphere decoder (SD) technique. The successive interference methods outperform the ZF and MMSE however their complexity is higher due to iterative nature of the algorithms.

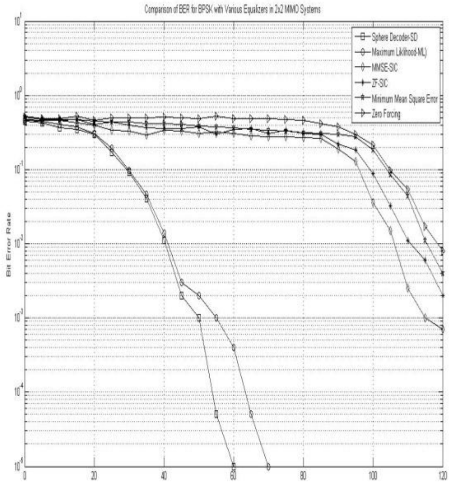


Figure 9.

"Fig.9"shows the comparison of BER for BPSK with different equalizers.

VI. CONCLUSIONS

When equalization is done through frequency selective fading channel, then the performance of ZF, MMSE, ZF-SIC and MMSE-SIC are very poor. ML provides the better performance in comparison to other equalizers. Sphere decoder provides the best performance and the highest decoding complexity as compare to ML. BER performance in order to highest to lowest is as: SD > ML > MMSE-SIC > ZF-SIC > MMSE > ZF. We finally concluded that SD is best suited method to remove ISI in Frequency selective fading channel in MIMO systems.

REFERENCES

[1] [1] H. El Gamal and A.R. Hammons, "The layered space-time architecture: a new perspective", IEEE Trans. Inform. Theory, vol. 47, pp. 2321–2334, Sept. 2001.

[2] Simon, M. K. and Alouini, M. Digital Communication over Fading Channels. John Wiley & Sons, 2004.

[3] G.J. Foschini and M.J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas", Wireless Personal Communications, vol. 6, pp. 311–335, 1998.

[4] G.J. Foschini , "Layered space-time architecture for wireless communications in a fading environment when using multiple antennas", Bell Labs. Tech. J., vol. 6, no. 2, pp. 41–59, 1996.

[5] M. Janakiraman ("Space-time codes and MIMO systems", Artech House, 2004.

[6] J .C. Liberti and T. S. Rappaport "Smart Antennas for Wireless Communications, Prentice H all PTR, 1999.

[7] Wang and G. B. Giannakis, "A simple and general parameterization quantifying performance in fading channels," IEEE Trans. Communication., vol. 51, no. 8, pp. 1389–1398, 2003.

[8] J. G. Proakis, Digital Communications, McGraw-Hill series in electrical and computer engineering, 1995.

[9] Ezio Biglieri, Robert Calderbank, Anthony Constantinides, Andrea Goldsmith, Arogyaswami Paulraj, H.Vincent Poor "MIMO Wireless Communications", Cambridge University Press, (2007).

[10] G. Arslan, B. L. Evans, and S. Kiaei , "Equalization for Discrete Multitone Receivers To Maximize Channel Capacity", IEEE Transactions on Signal Processing, submitted March 30, 2000.

[11] D. Shiu P. J. Smith D. Gesbert, M. Shafi and A. Nayguib, "From theory to practice: An overview of MIMO space–time coded wireless systems," IEEE J. Select. Areas Commun. vol., 21, no. 3, pp. 281–302, 2003.

[12] R. U. Nabar A. J. Paulraj, D. A. Gore and H. Bolcskei, "An overview of MIMO communications—a key to gigabit wireless," Proceedings of the IEEE, vol. 92, no. 2, pp. 198–218, 2004.

[13] G.Ginis and J.M.Cioffi, "On the relationship between V-BLAST and GDFE," IEEE Communications letters, vol. 5, pp. 364-366, 2001.

[14] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental trade-off in multiple-antenna channels," IEEE Transactions on Information Theory, vol. 49, pp. 1073-1096, May 2003.

[15] G. D. Golden , G. J. Foschini, R. A. Valenzuela and P. W. Wolniansky, " Detection algorithm and initial laboratory results using v-blast space-time communication architecture,"IEEE Electronic Letters, Vol.35, No.1, pp.14~16, January1999.

[16] J. Zhang, T. Bhatt, G. Mandyam "Efficient linear equalization for high data rate downlink CDMA signaling", 37th IEEE Asilomar Conference on Signals, Systems and Computers, 2003.

[17] V. K. Garg and J .E. Wilkes, "Wireless and Personal Communications Systems' Prentice Hall, 1996.

[18] Gesbert "MIMO space–time coded wireless systems,"presentation available at http://www.tele.ntnu.no/projects/beats/course.htm. Sept. 2003.

[19] D. Shiu and J. M. Kahn "Layered space-time codes for wireless communications using multiple transmit antennas," Vancouver, Canada, 1999.

[20] G. Foschini, G. Golden, R. Valenzuela and P. Wolniansky "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays", IEEE Journal on Selected Areas in Communications , vol. 17, pp. 1841–1852, 2000.

[21] D. Wubben, R. Bohnke, J. Rinas, V. Kuhn and K .D. Kammeyer "Efficient algorithm for decoding layered s pace-time codes", Electronics Letters, vol. 37, pp. 1348–1350, 2001.

[22] K.Lo, S.Marinkovic, Z C hen and B.Vucetic "BER performance comparison of layered space time codes", New York, USA ICC 2002.

[23] Choo,Y,S, ,Kim,J.,Yang,W.Y., and Kang C.G "Mimo-OFDM Wireless communication with matlab", IEEE PRESS, John Wiley and sons (Asia) Pte Ltd.

[24] C. E. Proakis, "Digital Communications," McGraw-Hill International Editions, New York, 4th edition, 2000.

[25] H. Jafarkhani "Space-time coding: Theory & Practice", Cambridge University Press, 2005.

[26] I.E. Telatar "Capacity of multi-antenna Gaussian channels, "European Transactions on Telecommunications, vol. 10, no.6, pp.585-595, 1999.

[27] S.Loyka and F. Gagon "Performance analysis of the V-BLAST algorithm: an analytical approach, "IEEE Transactions on Wireless Communications. Vol. 3, pp. 1326-1337, July 2004.

[28] K.I.Pedersen, J.B.Anderson, J.P.Kermoal and P.E.Mogensen "A stochastic multiple-input multiple-output radio channel model for evaluation of space-time coding algorithms," in Proc. VTC 200 Fall, Boston, vol. 2, pp.893-897, Sep. 2000.

[29] M.Varanasi and T.Guess "Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel," Conference Record of the Thirty-First Asilomar Conference on signals, Systems and computers, vol. 2, pp. 1405-1409, Nov-2-5 1997.

[30] A.M.Tulino and S.Verdu "Random Matrix Theory and Wireless Communications, Hanover MA 02339, USA: now publishers Inc., 2004.

[31] E.Biglieri, J.Proakies and S.Shamai 'Fading Channel Information Theoretic and Communication Aspects", IEEE Trans. On information Theory, vol. 44, pp.2619-2692, Oct. 1998.

[32] X.Li, H.Huang, G.J.Foschini, and R.A.Valenzu "Effects of Iterative Detection and Decoding on the Performance of BLAST", IEEE Global Telecommunications Conference, vol.2, pp.1061-10066, Nov 2000.

[33] David Tse and Pramod Viswanath," Fundamentals of Wireless Communication Cambridge University Press, 2005.

[34] John G. Proakis and Masoud Salehi "Contemporary Communication Systems using Matlab," Brooks/Cole, 2000.

[35] Rohit Gupta, Amit Grover," BER Performance Analysis of MIMO Systems Using Equalization Techniques",Innovative System Design and Engineering, vol.3, no.10, pp. 11-25, 2012.

AUTHORS PROFILE

Amit Grover(M'06-SM'09- PI'11&12) The author became a Member (M) of Association ISTE in 2006, a Senior Member (SM) of society SELCOME in September 2009, and a Project-In charge (PI) in august 2011 and in September 2012. The author place of birth is Ferozepur, Punjab, India on 27th, September 1980.

The author received M. Tech degree in Electronics and Communication Engineering from Punjab Technical University, Kapurthla, Punjab, India in 2008 and received B. Tech degree in Electronics and Communication Engineering from Punjab Technical University, Kapurthala, Punjab, India in 2001. Currently, he is working as an Assistant Professor in Shaheed Bhagat Singh State Technical Campus, Ferozepur, Punjab. His area of interest includes signal processing, MIMO systems, wireless mobile communication, high speed digital communications and 4G wireless communications.

Test Case PrioritizationUsing Fuzzy Logic for GUI based Software

Neha Chaudhary,
IT Department,
JSS Academy of Technical Education, Noida, India

Om Prakash Sangwan,
School of ICT, Gautam Buddha University,
Greater Noida, India

Yogesh Singh,
Vice-Chancellor,
MS University, Baroda,
Gujarat, India

***Abstract*-Testing of GUI (Graphical User Interface) applications has many challenges due to its event driven nature and infinite input domain. It is very difficult for any programmer to test for each and every possible input. When test cases are generated using automated testing tool it uses each and every possible combination to generate test cases hence generates numerous number of test case for any GUI based application. Within a defined time frame it is not possible to test every test case, that is why test cases prioritization is required. Test-case prioritization has been widely proposed and used in recent years as it can improve the rate of fault detection during the testing phase. Very few methods are defined for GUI Test case prioritization that usually consider single criteria for assigning priority for the test case which is not sufficient for the consideration of that test case as more fault revealing. In this paper we have proposed a method for assigning weight value on the basis of multiple factors as one of the criteria for test case prioritization for GUI based software. These factors are: The type of event, Event Interaction, and Parameter-value interaction coverage-based criteria. In the proposed approach priority is assigned based upon these factors using fuzzy logic model. Experimental results indicate that the proposed model is suitable for prioritizing the test cases of GUI based software.**

Keywords-Graphical user Interface; Prioritization; Test Suite; Fuzzy Model.

I. INTRODUCTION

Testing is widely recognized as a key quality assurance (QA) activity in the software development process. Research in testing has received considerable attention in the last two decades [2,8,20,14]. Testing of graphical user interfaces (GUIs) was a neglected research area till last decade [4]. Graphical User Interface (GUI) constitutes as much as 45-60% of the total software code in any software, so testing of GUI is a very important concern [3,16]. Most of the test case generation techniques require human involvement and are resource intensive. Many automated approaches were proposed for test case generation but in practice capture replay tools are used [17]. So generation of test cases is a costly effort. Rapid prototyping model is followed for GUI development which involves continuous modifications in software versions [1]. Due to event driven nature of GUI it takes sequence of events as input and after change of state generates new sequence of input as output [6, 12, 19]. For different set of state and combination of inputs GUI generate different output [5, 20]. It would be difficult to manage all the combinations for testing as number of combination grows exponentially with the number of events. Running all GUI test cases and then fixing all bugs may be time consuming and delaying the project completion. This would require that the test developed for one version should be reusable across various versions [1, 4]. It is important to prioritize the test cases that uncover the most faults as fast as possible in the testing process. So prioritization of test suite is a challenging area [7, 9,13,18]. In this paper multiple factors are considered for the assignment of weight value for test suite.

This paper is organized as follows: Section II describes the research background for the proposed work. Section III describes the factors affecting the fault detection capability of test suite. Section IV introduces the concept of the proposed fuzzy model. Section V discusses about the experimental design. The results are displayed in section VI and conclusion and future work is presented in section VII.

II. RESEARCH BACKGROUND

The significant work is done by Renee C. Bryce and Atif M. Memon for test suite prioritization by interaction coverage. Test suite for GUI based program is prioritized by t-way interaction coverage and rate of fault detection is compared with the fault detection by other prioritization criteria [9]. Experimental results shows that test suits with the highest event interaction coverage benefit the most and test suits that has less interaction coverage does not benefit in using this prioritization technique.

In this approach only event interaction coverage criterion is taken as a measure for prioritizing test cases, there could possibly be significant effect of type of event in a test case, which will affect the rate of fault detection.

Atif M. Memon & Renee C Bryce provided a single abstract model for GUI and web application testing. In this approach test cases are prioritized by set of count based criteria, set of usage-based frequency and set of interaction based criteria [10]. The results show that test case prioritization by 2-way (interaction based criteria) and PV-LtoS (Parameter count based criteria) provided best improvement in the rate of fault detection for GUI based software. The main drawback of this technique is that the combination of different prioritization criteria is used and it is said that this is more effective than a single criterion.

However in order to cover web applications and GUI applications various factors need to be added and they add complexity to the process which can be avoided if specific criteria for web based application and GUI based application would be used.

In the work done by Chin-Yu Huang et al. on GUI Test case prioritization, weighted event flow graph was used for solving the non-weighted GUI test cases and ranked GUI test cases based on weight scores. In order to assign weights, events are classified based on their importance in the GUI application [11, 15]. In this technique weight summation of termination event and unrestricted focus event is equal to that of restricted focus event which requires further research in this area. In this approach the effect on fault detection based on event interaction with other event need to be explored further. Weight value of each interaction would also have impact on fault detection ability of test cases which was not considered in this approach.

III. Factors For Test Case Prioritization

Weight value will be assigned by considering following factors:

- Type of event
- Event Interaction
- Count based criteria

In following section we will elaborate different criteria considered for assigning weight values:

- Type of event

Type of event, a test suite is covering has significant impact on the fault revealing capability of test case.

According to the literature survey events are classified as following five types, restricted-focus event, unrestricted-focus event, termination event, menu-open event, and system-interaction event. Event weight has been assigned on the basis of importance of specific type of events [15]. This categorization of events is given in table 1.

TABLE 1: EVENT WEIGHT ASSIGNMENTS

Event type	Weight Value
Restricted-focus event	5
System-interaction event	4
Termination event	3
Menu-open event	2
Unrestricted-focus event	1

- Event Interaction

In event driven software event interaction makes the program to follow a different execution path that may reveal faults in the system. In our proposed method Priority is assigned to those test cases which have large number of parameter value interaction [9].

- Count-based criteria

Since the GUI is the collection of events, number of actions performed with events, set of parameters and number of windows. It is very important that test suit that provides maximum count coverage should be given higher importance then the test suit that provide low coverage. So another factor that will be considered is the count of number of windows, actions or parameter values that a test case may cover.

IV. Proposed Fuzzy Model

Fuzzy logic is a convenient way to map an input space to output space. In this paper we have proposed a fuzzy model with three inputs, namely Type of event, Event Interaction, Count based criteria. Figure1 shows the fuzzy model. The proposed model consists of three inputs and provides a crisp value of priority using Rule Base.

Fuzzy Inference System (FIS) is the process of formulating the mapping from a given input to an output using fuzzy logic. This will use Mamdani’s fuzzy inference method which is most commonly seen fuzzy methodology as shown in Figure 2.

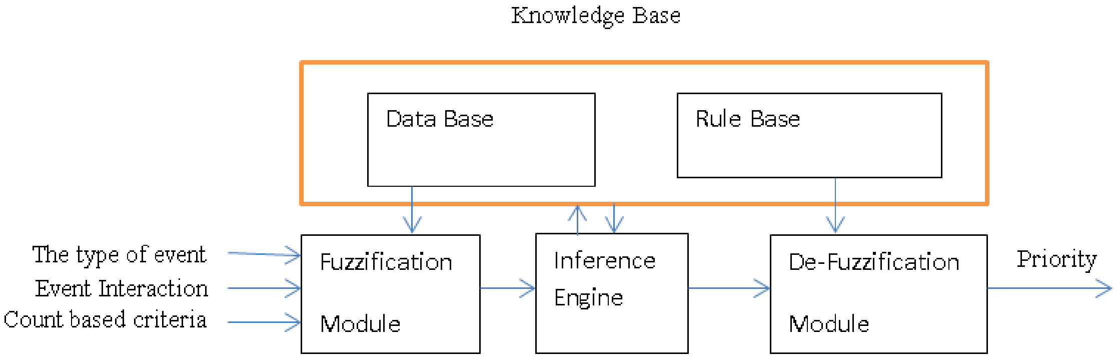


Figure 1: Fuzzy Model for Prioritization

After the fuzzification process, there is a fuzzy set for each output variable that needs defuzzification. The input for the defuzzification process is a fuzzy set (the aggregate output fuzzy set) and the output is singleton number. Further centroid method will be used for defuzzification. Centroid method will return the centre of area under the curve.

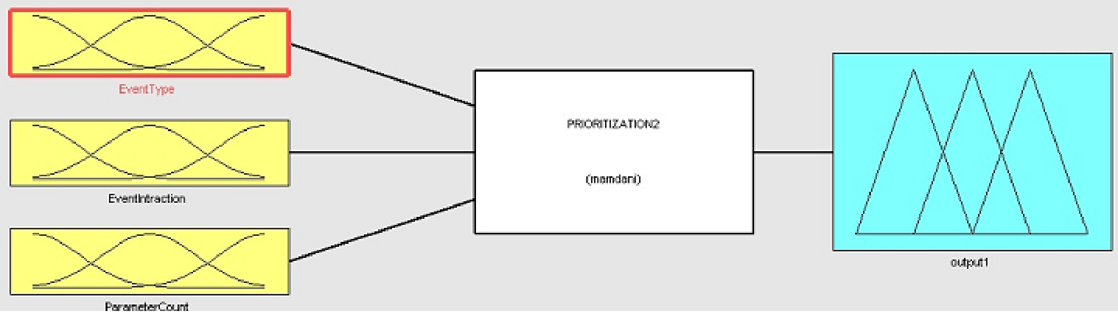


Figure 2: Fuzzy Inference System: Priority Model

V. Experimental Design

In order to fuzzify the inputs, we have selected following membership functions for the Type of event, Event Interaction and Count based criteria and they are shown in figure 3-5. GUI events are classified into five categories and they have different fault revealing capabilities. Weight value of test case will be calculated by taking summation of weight according to categorization. That weight will be divided into five states (linguistic variables) i.e. very low, low, medium, high and very high as shown in figure 3. The input variable Event Interaction has been divided into five levels i.e. very low, low, medium, high and very high as shown in figure 4.

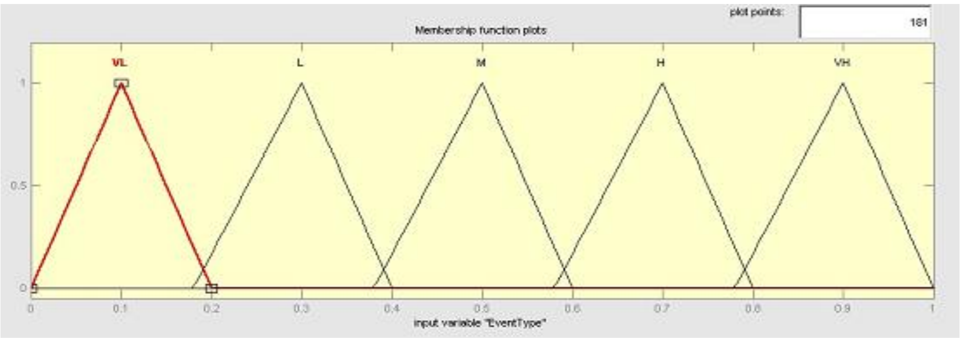


Figure 3: Fuzzification of Input Variable Event Type

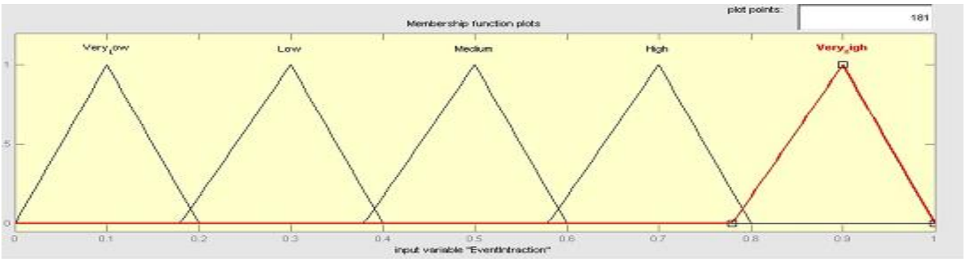


Figure 4: Fuzzification of Input Variable Event Interaction

Similarly the input variable count has been divided into five states i.e. very low, low, medium, high and very high as shown in figure 5.

The output variable priority is classified as very low, low, medium, high and very high. Similarly priority has five membership functions as shown in figure 6:

A. Rule Base and Evaluation Process

When input data is fuzzified, processing is carried out in fuzzy domain. The model integrates the effects of multiple factors type of event, Event Interaction and Count based criteria into a single measurable parameter that will define the priority of test case, based on the following knowledge/rule base. The rule base can further be advanced by creating more ranges (fuzzy sets) for the input variables. All inputs and outputs are fuzzified as shown in figure 3 to 6. All possible combinations of inputs were considered that will create 5^3 i.e. 125 sets. The priority for all 125 combinations is classified as very low, low, medium, high & very high by expert judgment. This indicates to formulation of 125 rules for the fuzzy model and some of the rules are presented below:

1. If value assigned for Type of event is low, Event Interaction is low and Count based is low then priority will be low.

2. If value assigned for Type of event is medium, Event Interaction is medium and Count based is medium then priority will be medium.

3. If value assigned for Type of event is low, Event Interaction is high and Count based is high then priority will be medium.

.

.

.

- If value assigned for Type of event high, Event Interaction is medium and Count based is high then priority will be high.

…..

125. If value assigned for Type of event very low, Event Interaction is high and Count based is very low then priority will be low.

All 125 rules are inserted and rule base is created. A rule is fired based on the particular set of inputs. In this model Mamdani style inference has been used.

The output of test case priority has been observed using rule viewer for particular set of inputs using MATLAB fuzzy Tool Box as shown in figure 7.

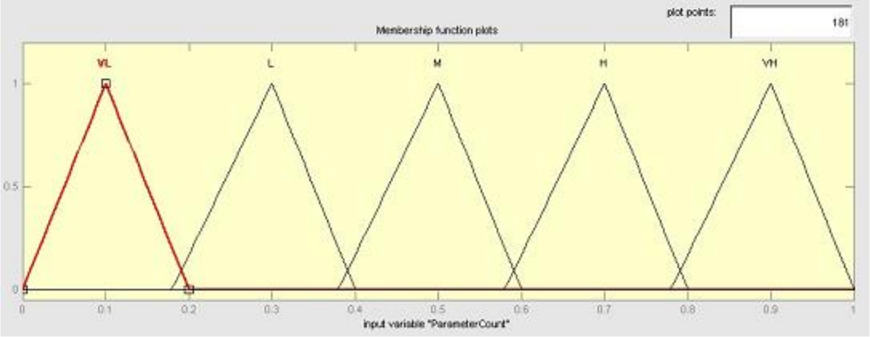


Figure 5: Fuzzification of Input Variable Count

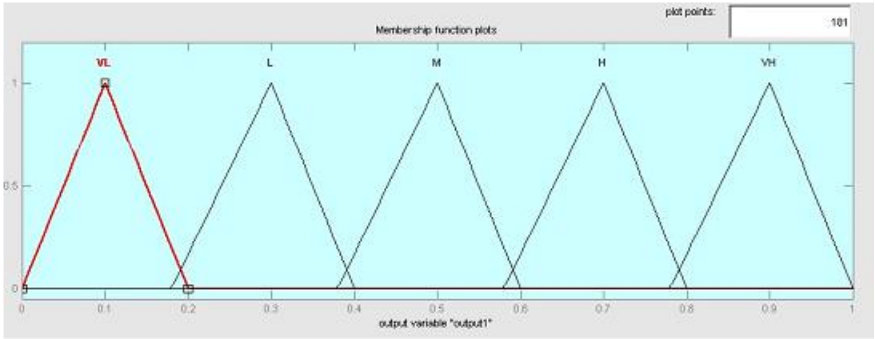


Figure 6: Fuzzification of Input Variable Priority

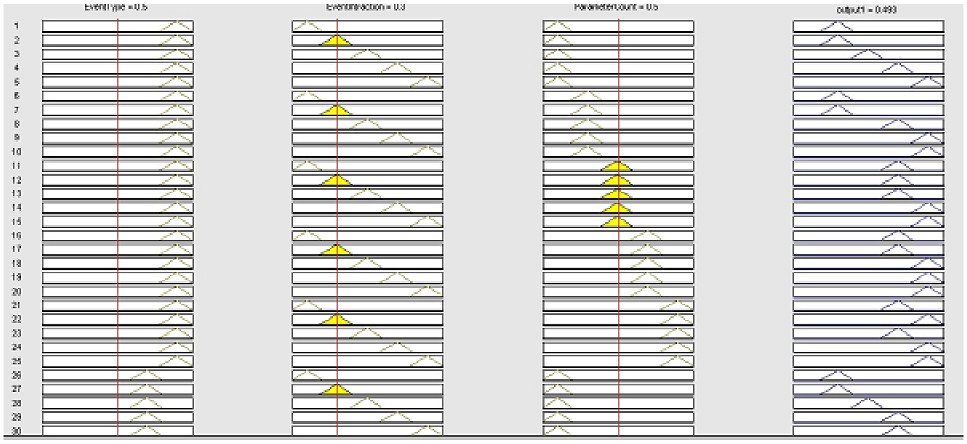


Figure 7: Rule Viewer for the Priority Model

VI. Experimental Results

For example we have following crisp value inputs to the model: type of event =0.5, Event Interaction =0.3 and Count based criteria=0.5.

These inputs are provided for the fuzzification module and after fuzzification of given value we find the type of event =0.5 belongs to the fuzzy set low with membership grade 0.9 and belongs to fuzzy set medium with membership grade 0.9 and with high it has membership grade 0.72.

For event Interaction =0.3 belongs to the fuzzy set low with membership grade 0.9 and belongs to fuzzy set medium with membership grade 0.72 and with high it has membership grade 0.72. Count based criteria=0.5 belongs to the fuzzy set low with membership grade 0.9 and belongs to fuzzy set medium with membership grade 0.9 and with high it has membership grade 0.72. With these input values we find that rules given in table 2 will be considered:

First rule assigns the priority low to an extent of 0.9 and second rule gives priority medium to an extent of 0.72 and the third rule gives priority high to an extent 0.72 this is shown in the figure 8.

TABLE 2: TEST SUITE PRIORITY CALCULATION FOR A GIVEN INPUT SET

The type of event (.5)	Event Interaction(.3)	Count based criteria(.5)	Priority	Membership Grade for Test Case Priority
Low	Low	Low	Low	Min(0.9,0.9,0.9)=0.9
Medium	Medium	Medium	Medium	Min(0.9,0.72,0.9)=0.72
High	Medium	High	High	Min(0.72,0.72,0.72) =0.72

A. Defuzzification

After getting the fuzzified output as specified in previous section, we defuzzify them to get the crisp value of the output variable priority [21]. Transformation of the output from fuzzy domain to crisp domain is called defuzzification. In this model we defuzzify using centre of gravity (COG) method of the aggregate output 1, 2 and 20. X axis centroid points for all three variables are 2.9, 4.9 & 6.9 the final value for GOG is 4.84.

The effect of these rules is also observed by simulating the model using fuzzy logic tool box of MATLAB. The priority for the given input values comes out to be 0.493 which is the same as calculated from COG method.

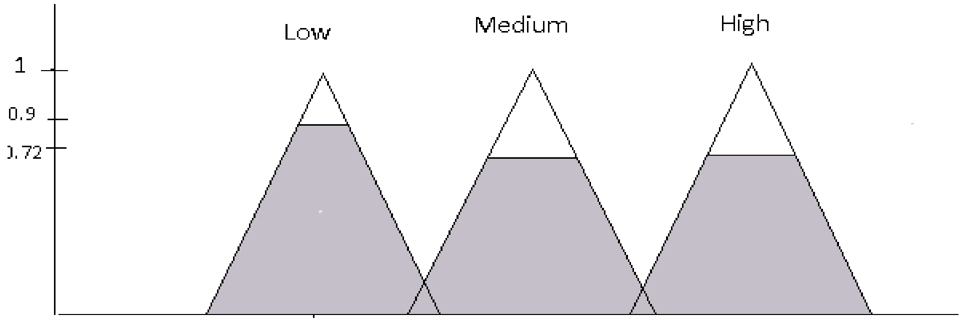


Figure 8: Output computation for Test Case Priority

VII. Conclusion

The problems of test-case prioritization have been explored in this paper to improve the rate of fault detection effectiveness for GUI based software. We have proposed a fuzzy based technique to assign priority of test case. Priority of test case will be assigned as very low, low, medium, high and very high. In this technique three factors namely Type of event, Event Interaction, Count based criteria are considered to assign weight values for test cases. Impact of these factors are categorize in five categories as very low, low, medium, high and very high. Experimental results shows that the proposed fuzzy model is proved to be an effective approach for test case prioritization for GUI based software.

References

[1] A. M. Atif, “Automatically repairing event sequence based GUI test suites for regression testing”, ACM Transaction on Software Engineering and Method. Volume 18, Issue 2, Nov. 2008.

[2] M. J. Harrold. “Testing: a roadmap”. In ICSE '00: Proceedings of the Conference on The Future of Software Engineering, ACM Press, New York, NY, USA, pages 61.72, 2000.

[3] B. A. Myers. User interface software tools. ACM Transactions on Computer-Human Interaction,1995.

[4] A. M. Memon. A Comprehensive framework for testing graphical user Interfaces. Ph.D. thesis, Department of Computer Science, University of Pittsburgh, July 2001.

[5] L. White and H. Almezen, "Generating test cases for GUI responsibilities using complete interaction sequences". In Proceedings of the International Symposium on Software Reliability Engineering, pages 110-121, Oct. 2000.

[6] R. K. Shehady and D. P. Siewiorek, "A method to automate user interface testing using variable finite state machines." In Proceedings of The Twenty-Seventh Annual International Symposium on Fault-Tolerant Computing (FTCS'97), Washington - Brussels - Tokyo, pages 80.88, June 1997.

[7] Elbaum S., Malishevsky A. G., Rothermel G.,"Test case prioritization: a family of empirical studies", IEEE Transactions on Software Engineering vol. 28 (2), pp. 159–182, 2002.

[8] A. M. Atif ,Mary Lou Soffa, Martha E. Pollack," Coverage criteria for GUI testing" Proceedings of the 8th European Software Engineering conference held jointly with 9th ACM SIGSOFT international symposium on Foundations of Software Engineering, pp. 256-267, 2001.

[9] Renee C. Bryce, Atif M Memon, "Test suite prioritization by interaction coverage" Domain-Specific Approaches to Software Test Automation Workshop September, 2007.

[10]Renee C. Bryce, Sreedevi Sampath, Atif M Memon, "Developing a single model and test prioritization station for event- driven software" IEEE Transaction on Software Engineering, Jan 2010

[11]Chin-Yu Huang, Jun-Ru Chang, Yung-Hsin Chang, "Design and analysis of GUI test-case prioritization using weight-based methods" The journal of Systems and Software 83, pp 646-659, 2010

[12]A. M. Atif ,Mary Lou Soffa, Martha E. Pollack," coverage criteria for GUI testing" Proceedings of the 8th European Software Engineering conference held jointly with 9th ACM SIGSOFT international symposium on Foundations of Software Engineering, pp. 256-267, 2001.

[13]Scott McMaster, Atif M. Memon,"Call-Stack coverage for GUI test suite reduction" IEEE Transaction on Software Engineering, Volume 34, Jan/Feb, 2008

[14]Jaymie Strecker , Atif M Memon ,"Relationships between test suites, faults, and fault detection in GUI testing" In ICST '08 Proceedings of the First international conference on Software Testing, Verification, and Validation, (Washington, DC, USA), 2008.

[15]Paul Gerrard, "Testing GUI applications", EuroSTAR, Edinburgh UK, 1997.

[16]A. Memon, A. Nagarajan, and Q. Xie, "Automating regression testing for evolving GUI software," J. Software Maintenance and Evolution: Research and Practice, Volume 17, no. 1,pp. 27-64, 2005.

[17]Atif M. Memon , Qing Xie "Studying the fault-detection effectiveness of GUI test cases for rapidly evolving software" IEEE Transaction on Software Engineering, Volume 31, no. 10, pp. 884-896, Oct. 2005

[18]S. McMaster and A. Memon, "Call Stack Coverage for GUI test-suite reduction", Proc., 17th International Symposium on Software Reliability Engineering, Nov. 2006.

[19] A. M. Memon, M. E. Pollack, and M. L. Soffa, "Hierarchical GUI test case generation using automated planning", IEEE Transactions on Software Engineering, pp 144–155, Feb. 2001.

[20] Daniel R. Hackner , Atif M. Memon ,"Test case generator for GUITAR" , International Conference on Software Engineering, (Washington, DC, USA), 2008.

[21]Yogesh Singh, Pradeep Kumar Bhatia, Omprakash Sangwan,"Predicting software maintenance using fuzzy model", published in SIGSOFT Software Engineering Notes, vol 34, July 2009.

Masking Digital Image using a Novel technique based on a Transmission Chaotic System and SPIHT Coding Algorithm

Masking Digital Image

[1]Hamiche Hamid
Laboratoire de Conception et Conduite des Systèmes
de Production, UMMTO, BP 17 RP, 15000,
Tizi-Ouzou, Algérie

[2]Lahdir Mourad
Laboratoire d'Analyse et Modélisation des Phénomènes
Aléatoires, UMMTO, BP 17 RP, 15000,
Tizi-Ouzou, Algérie

[3]Tahanout Mohammed
Laboratoire d'Analyse et Modélisation des Phénomènes
Aléatoires, UMMTO, BP 17 RP, 15000,
Tizi-Ouzou, Algérie

[4]Djennoune Said
Laboratoire de Conception et Conduite des Systèmes
de Production, UMMTO, BP 17 RP, 15000,
Tizi-Ouzou, Algérie

Abstract— **In this article, a new transmission system of encrypted image based on novel chaotic system and SPIHT technique is proposed. This chaotic system is made up of two chaotic systems already developed: the discrete-time modified Henon chaotic system and the continuous-time Colpitts one. The transmission system is designed to take profit of two advantages. The first is the use of a robust and standard algorithm (SPIHT) which is appropriate to the digital transmission. The second is to introduce farther complexity of the encryption using the chaotic system over secure channel. Through these two advantages, our purpose is to obtain a robust system against pirate attacks. Cryptanalysis and various experiments have been carried out and the results were reported in this paper, which demonstrate the feasibility and flexibility of the proposed scheme.**

Keywords- Chaos Modified Henon;Colpitts, SPIHT; Robustness.

I. Introduction

The fascinating developments in digital image processing and network communications during the past decade have created a great demand for real-time secure image transmission over the Internet and through wireless networks. To meet this challenge, a variety of encryption schemes have been proposed [1-4]. Nevertheless, conventional image encryption algorithm such as data encryption standard (DES) is not suitable for image encryption. Because of the special storage characteristics of an image [5] and weakness of low-level efficiency when the image is large [6-7]. The chaos-based encryption has suggested a new and efficient way to deal with the intractable problem of fast and highly secure image encryption.

Chaotic systems have many important properties, such as the sensitive dependence on initial conditions, system parameters, the density of the set of all periodic points and topological transitivity. Most properties are related to some requirements such as mixing and diffusion in the sense of cryptography [8]. Matthews proposed a chaotic encryption algorithm in 1989 [9]. Since then, researches of image encryption technology which are based on chaotic systems are increased [10-14]. These methods have high-level efficiency but also weakness, such as small key space, weak security and complexity to overcome these drawbacks. In this paper, a novel image encryption scheme incorporating a chaotic map is introduced. This scheme integrates chaotic encryption into the process of bit stream generation by an SPIHT (Set Partitioning In Hierarchical Tree) encoder. The proposed scheme only introduces few overheads to the image coder by using selective encryption, i.e., only sensitive bits in the compressed stream are encrypted.

Meanwhile, since a cipher-text feedback mechanism is employed, many powerful attacks such as the known-plain-text attack are not valid to break the designed cryptosystem. Furthermore, due to the use of chaotic pseudo-random bits, which efficiently masks the SPIHT coding bitstream, the ciphered stream can be truncated at any position while keeping the obtained bits decipherable. The rest of the paper is organized as follow. Section 2 discuses the proposed chaos-based image encryption scheme. Section 3 shows some numerical results. In section 4, we analyze the security of the new chaotic encryption scheme. Finally, section 5 concludes the paper.

II. The Proposed Chaos-Based Image Encryption Scheme

In this work a communication system based on chaos encryption and SPIHT coding is realized. The global scheme of the proposed system for private communications is shown in Fig. 1. Note that the transmission channel is a public one. Consequently, any hacker has a free access to information passing through the channel which is considered perfect in our works.

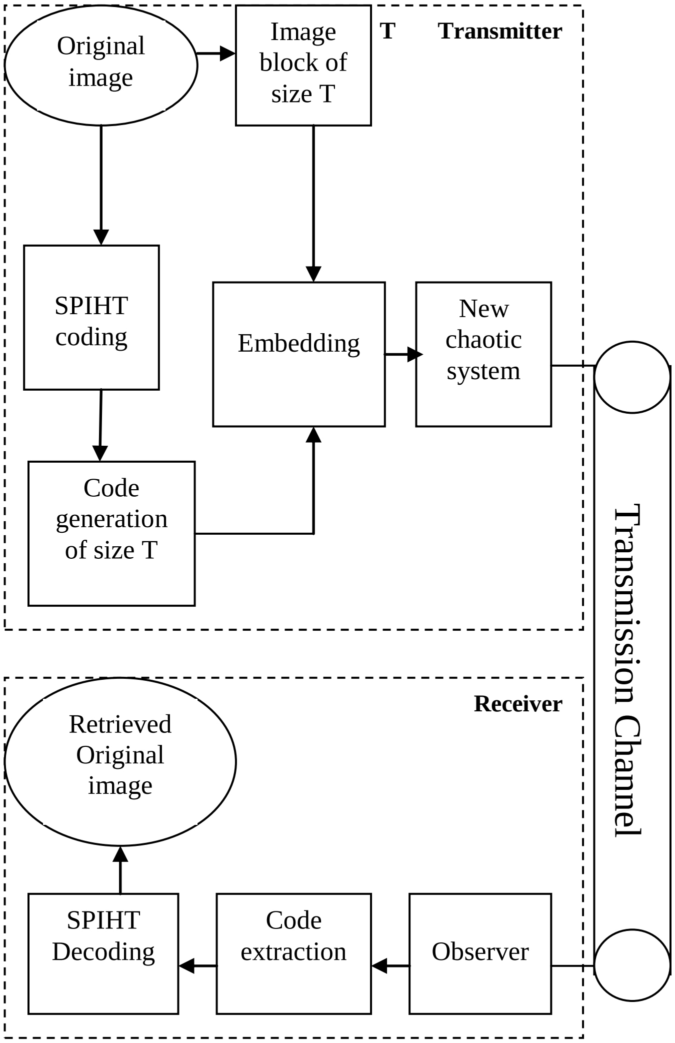


Figure 1. Transmission chain based on a chaotic dynamical system and SPIHT coding

III. Transmitter Presentation

The transmitter is composed of two main blocks (see Fig. 1): a SPIHT coding block and chaotic system block.

A. SPIHT Coding

Progressive coding (also called embedding coding) refers to the way that the most significant bits representing an image are placed at the beginning of the code. The code bits are arranged according to their importance relative to the representation of the image. A decoder can truncate the code at any position and obtain an estimate of the image based on the information up to that particular point. There are two well-known progressive coding schemes which are EZW (Embedded Zerotree Wavelet coding) [15] and SPIHT algorithms [16-19]. Note that SPIHT algorithm is more efficient than EZW. After the subband decomposition is applied to the concerned image, the SPIHT algorithm works by partitioning the subband-decomposed image into significant and insignificant partitions by using the following function:

$$S_n(T) = \left. \begin{matrix} 1 \max\limits_{i,j\in T} \left\{\left|C_{i,j}\right|\right\} \geq 2^n \\ 0 \text{ otherwise} \end{matrix} \right\} \qquad (1)$$

Where $S_n(T)$ represents the significance of a set of coordinates T and $C_{\{i,j\}}$ is the coefficient value at coordinate (i, j).

There are two passes in the algorithm: the sorting pass and the refinement pass. Three lists are defined, which are the list of insignificant sets (LIS), list of insignificant pixels (LIP), and list of significant pixels (LSP), respectively. The LIP and LSP consist of nodes that contain single pixels, while the LIS contains nodes that have descendants. The sorting pass is performed on these three lists and finally makes pixels in LSP, which is arranged in an order according to the information importance. The maximum number of bits required to represent the largest coefficient in the spatially oriented tree is designated as n_{max}, computed by the following formula:

$$n_{\max} = \left\lfloor \log_2 \left(\max_{i,j} \left\{ \left| C_{i,j} \right| \right\} \right) \right\rfloor \qquad (2)$$

During the sorting pass, those coordinates of the pixels that remain in LIP are tested for the significance. The result $S_n(T)$ is then sent to the output. Those that are significant will be moved to LSP, along with their sign bits output. Sets in LIS will also have their significance tested and, if they are found to be significant, they will be removed. Consequently the result will be partitioned into subsets. Subsets with a single coefficient, if found to be significant, will be added to LSP, or else they will be added to LIP.

During the refinement pass, the *nth* most significant bit of the coefficients in LSP is an output. The value of n is then decreased by 1 and the sorting and refinement passes are repeated.

This process continues until either the desired rate is reached, or $n = 0$, and all the nodes in LSP have their bits output. The latter case will result in an almost perfect reconstruction since all the coefficients have been processed completely.

There are two features reside in the SPIHT, which makes the design to be introduced in the next section. First, the SPIHT is not noise tolerant, i.e., the method is sensitive to small modification of bits in their bitstreams. Secondly, from the above discussion, it is clear that there are two kinds of data contained in a SPIHT-coded bitstream: they are named structure bits and data bits, respectively. Structure bits refer to those used for synchronizing the encoding end and the decoding end in the construction of spatially oriented tree. These bits are extremely sensitive to noise, especially the first few bits in the bitstream. Data bits refer to those coding signs of image coefficients or coding values of coefficients generated in the refinement pass. Change of data bits does not seriously affect the reconstruction of the image, but only introduces a small amount of noise to the result.

Thus, to protect an image, an efficient method is to encrypt only those structure bits.

B. New Chaotic System

The new chaotic system is the discrete-time modified Henon chaotic system which is coupled with the sampled continuous-time Colpitts chaotic system. The new chaotic system block is detailed as follows:

1) Discrete-time chaotic system

The discrete-time chaotic system is the modified Henon's map (see for example [20-21]). A simplified version of our proposed discrete scheme is:

$$\left.\begin{aligned} x_1(k+1) &= a - x_2^2\ (k) - bx_3(k) \\ x_2(k+1) &= x_1(k) \\ x_3(k+1) &= x_2(k) \\ y(k) &= x_2(k) \end{aligned}\right\} \quad (3)$$

where $x = [x_1\ x_2\ x_3]^T \in R^3$ denote the state vector and y(k) the output. Chaotic behavior of system (3) as shown by Fig. 2 is obtained by setting its parameters a=1.76 and b=0.1. These parameters are chosen such that system (3) exhibits chaotic behavior. Initial conditions $x_1(0)=1$, $x_2(0)=0.1$ and $x_3(0)=0.1$ are chosen inside the strange attractor basin.

In private communication, one of the main purposes is to increase the security. It is interesting to modify the system (3) by introducing in its dynamic, the sampled states of a continuous-time chaotic system. The continuous-time chaotic system used in our work is given as follow:

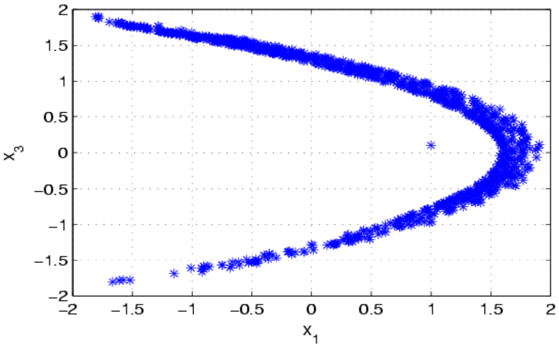


Figure 2. Chaotic attractor of modified Henon system

2) Continuous-time chaotic system

This system has been widely studied in the literature [22-23]. The state equations of the normalized Colpitts's oscillator in a continuous time are given as:

$$\left.\begin{aligned} \dot{z}_1 &= a_1(-\exp(-z_2)+1+z_3) \\ \dot{z}_2 &= a_2 z_3 \\ \dot{z}_3 &= -a_3(z_1+z_2) - a_4 z_3 \end{aligned}\right\} \quad (4)$$

Where $z = [z_1\ z_2\ z_3]^T \in R^3$ is the state vector and $a_1 = \frac{g}{q(1-k)}, a_2 = \frac{g}{qk}, a_3 = \frac{qk(1-k)}{g}, a_4 = \frac{1}{q}$.

To have a chaotic behavior as shown by Fig. 3, the parameters of system (4) are given as follows: g=4.46; q=1.38 and k=0.5. Initial conditions $z_1(0)=1.6$, $z_2(0)=8$ and $z_3(0)=0.1$ are chosen inside the strange attractor basin.

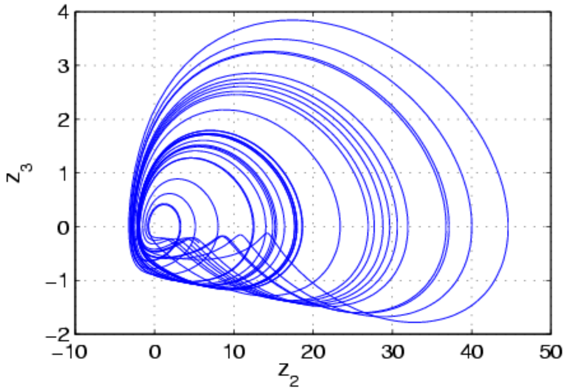


Figure 3. Chaotic attractor of Colpitts system

In order to make the transmitter more complex thus more robust (see Section 4), we have chosen to add the three states (z_1, z_2, z_3) and the message m to the third dynamic of the system (3). Then, the new obtained coupled system is:

$$\left.\begin{aligned} x_1(k+1) &= a - x_2^2\ (k) - bx_3(k) \\ x_2(k+1) &= x_1(k) \\ x_3(k+1) &= x_2(k) + A_1 z_1(nT) \\ &\quad + A_2 z_2(nT) + A_3 z_3(nT) + cm(k) \\ y(k) &= x_2(k) \end{aligned}\right\} \quad (5)$$

Where A_1, A_2, A_3 and c are the new parameters of the new discrete-time chaotic system, $n \in N$ and T is the sampling period of the system (4). To preserve the chaotic behavior of the system defined by (5), these parameters are chosen with precaution. In our case, we must respect the following values:

$A_1 \le 0.01, A_2 \le 0.01, A_3 \le 0.1$ and $c \le 1$. The strange attractor oh the new chaotic discrete-time is shown by the Fig. 5.

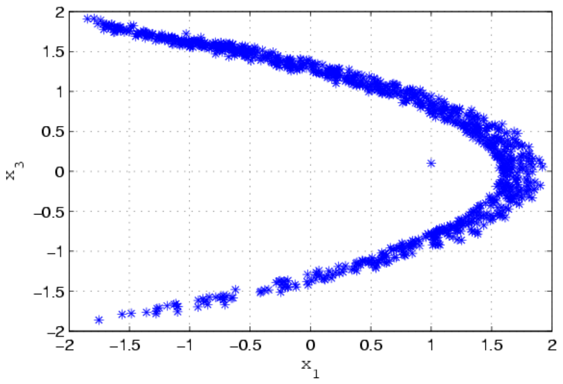


Figure 4. Chaotic attractor of new chaotic system

IV. Receiver Presentation

The receiver is also composed of two main blocks (see Fig. 1): a chaotic observer and SPIHT decoding block. The receiver blocks are detailed as follows:

A. Chaotic Observer

In this part, the system (5) with the output y(k)=x$_2$(k) is considered. For the reception, based on the works of [24-25], we have designed a delayed discrete observer that works with a sampling time T and which allows to reconstruct all states and the transmitted message m of (5). The design of the observer is detailed in the work [26], it is given as follow:

$$\left.\begin{array}{l} \hat{x}_1(k-1)=y(k) \\ \hat{x}_3(k-2)=\dfrac{a-y(k)-y^2(k-2)}{b} \\ \hat{m}(k-3)=\dfrac{a-y\ \ (k)-y^2(k-2)}{bc} \\ \quad -\dfrac{y(k-3)+A_1z_1(nT-3)+A_2z_3(nT\ \ -3)}{c} \\ \quad -\dfrac{A_3z_3(nT-3)}{c} \end{array}\right\} \quad (6)$$

B. SPIHT Decoding

The SPIHT decoding is the inverse process of SPIHT coding. So we may refer to the concrete procedures of the encryption algorithm as explained in the subsection 2.1.1. In the following section, the numerical results are given.

V. Numerical Results

Numerical results and performance analysis of the proposed image encryption scheme are provided in this section. Fig. 5 depicts the original image which is a 256×256 size 8 bits Lena image.

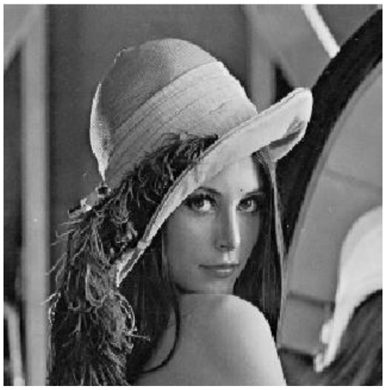


Figure 5. Original image

Figs. 6 and 7 show results of encrypting image and decrypting image, respectively. It can be seen that the decrypted image is clear and correct without any distortion with PSNR (Peak Signal to Noise Ratio) equal to 28.51 dB. This result shows clearly the performance of the used method.

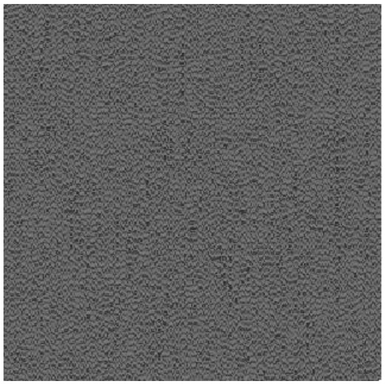


Figure 6. Encrypted image

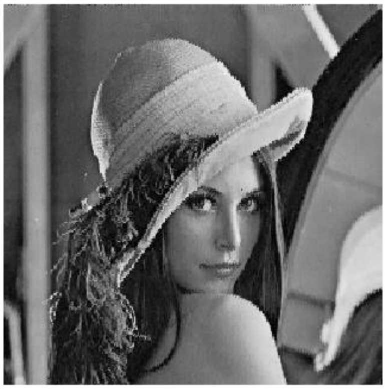


Figure 7. Decrypted image

Figs. 8 and 9 show some results of decrypting with 0.05 bpp (bits per pixel) and 1 bpp, respectively. In the following section, the robustness of the proposed scheme is studied.

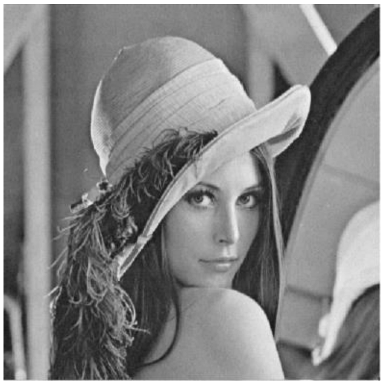


Figure 8. Decrypted image with 0.5 bpp

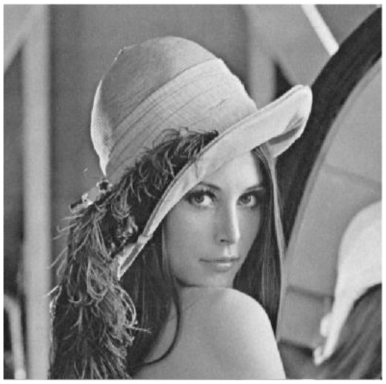


Figure 9. Decrypted image with 1 bpp

VI. Security Analysis

The security of the above-described chaos-based encryption scheme is now analyzed by studying two tests: histogram analysis and key space analysis.

A. Histogram Analysis

As expected, the test results show that the histogram of encipher-image is quasi uniform, which makes statistical attacks difficult.

This result is in accordance with the result given by Fig. 6, Figs. 10 and 11 show the histograms of plain and encrypted image, respectively.

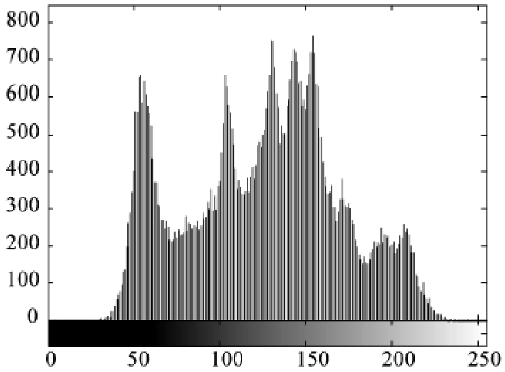


Figure 10. Histogram of plain image

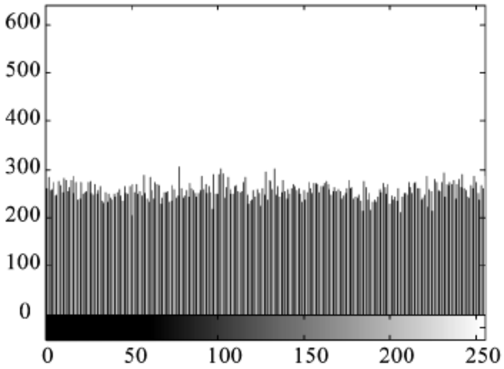


Figure 11. Histogram of encrypted image

B. Key Space Analysis

One of the most important properties of cryptographic systems is the existence of a secret key which defines the level of security of the cryptosystem [27-28]. The better secret key is designed, the larger is the key space hence and the more secure is the cryptosystem. Chaotic systems are well-known for their high sensitivity to initial conditions and parameters variations. From a cryptographical viewpoint, the initial conditions and the parameters of chaotic systems may be used to define a secret key for the chaos-based communication systems.

In the present case-study, in which we have used two chaotic systems in the transmitter (continuous and discrete systems), let us assume that the initial conditions are exactly known by a non-authorized intruder. We consider the parameters of the two systems to construct a secret key for our communication scheme. Firstly, we suppose that a non-authorized intruder knows the structure of the two chaotic systems without knowing exactly the true values of their parameters.

Let P_i:=(p_1=g, p_2=q, p_3=k, p_4=a, p_5=b, p_6=A_1, p_7=A_2, p_8=A_3) be the secret key. Our aim is to determine the size *r* of the key space K_s={P_1, P_2,..., P_r} which represents the finite set of all possible keys in order to evaluate the level of security produced by the secret key. To that end, we have to define the range of variation and the sensitivity of each parameter p_i, for i=1,...,8.

Without much loss of generality, we assume that the size s of the interval of variation of each parameter pi that leads to chaotic behaviors of the two systems is equal to 10^{-1}. Simulation experiments are carried out to evaluate the sensitivity S_i of each parameter P_i by determining the smallest parameter mismatch that gives us two different chaotic behaviors (i.e., two different attractors) when the rest of parameters p_j, for $j \in 1,2,...8/\, i$ are fixed.

Figs. 12 and 13 illustrate the sensitivity of the two systems to small changes of parameters g and a, respectively. The sensitivity to parameters is illustrated in TABLE I.

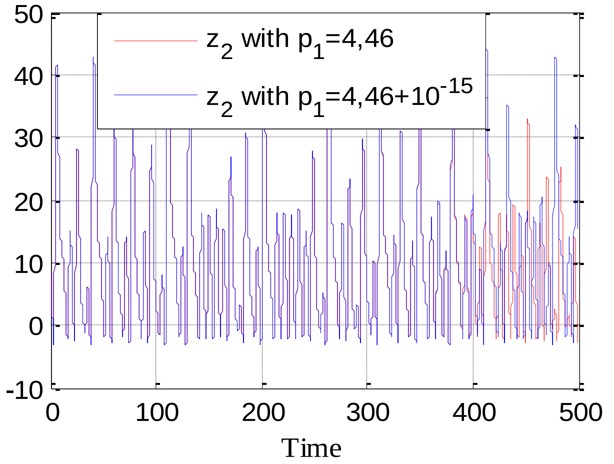


Figure 12. State z_2 of the continuous chaotic system for small changes (10^{-15}) of parameter g

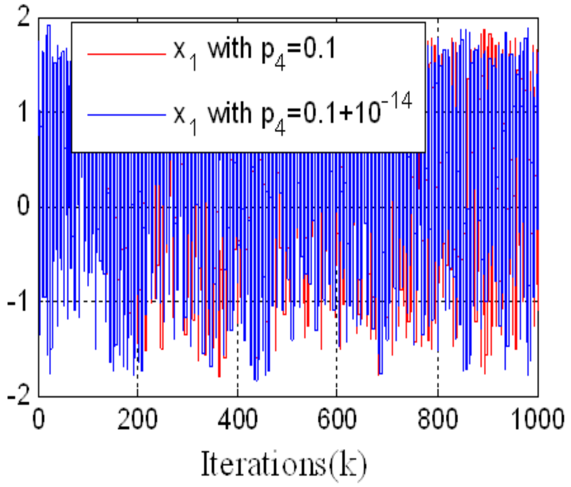


Figure 13. State z_3 of the new discrete chaotic system for small changes (10^{-14}) of parameter a

TABLE I. SENSITIVITY TO PARAMETERS

Chaotic system	**Parameters**	**Sensivity**	**Nb. of possibilities** $(N_i = s \times S_i^{-1})$
Colpitts chaotic system	p_1=g=4.46 p_2=q=1.38 p_3=k=0.5	$S_1=10^{-15}$ $S_2=10^{-15}$ $S_3=10^{-15}$	$N_1=10^{14}$ $N_2=10^{14}$ $N_3=10^{14}$
New chaotic system	p_4=a=1.76 p_5=b=0.1 $p_6=A_1$=0.01 $p_7=A_2$=0.01 $p_8=A_3$=0.1	$S_4=10^{-14}$ $S_5=10^{-14}$ $S_6=10^{-15}$ $S_7=10^{-15}$ $S_8=10^{-15}$	$N_4=10^{13}$ $N_5=10^{13}$ $N_6=10^{14}$ $N_7=10^{14}$ $N_8=10^{14}$

The size of the key space is:

$$r = \Pi_{i=1}^{8}(N_i) = 10^{14\times6+13\times2} = 10^{110}.$$

Relying on nowadays available computational power, a key space of size $O(2^{100})$ is generally required. In our case $r=10^{110} >> 2^{100}$ which means that the key space produced enhances a largely satisfactory level of security from a cryptographical viewpoint. The same reasoning may be applied to define and characterize a secret key for the first case-study.

Compared with other similar encryption schemes [10, 13-14, 17], our algorithm described above has higher security and can resist all kinds of known attacks, such as the known-plaintext attack and so on. Here, some security analysis results on the scheme are described, including the most important ones like keyspace analysis, statistical analysis, and differential analysis.

VII. CONCLUSION

In this paper, a novel chaotic image encryption scheme integrated with SPIHT wavelet image coding has been introduced. To overcome the drawbacks of small key space and weak obscure in the current chaotic encryption methods, its structural parameters are used as encryption key in chaotic. Experimental analysis demonstrates that the image encryption algorithm has the advantages of large key space and high-level security, high obscure level and high speed. Finally, numerical simulations results illustrate the effectiveness of the proposed method. To demonstrate the robustness of our system, further works are needed, particularly, the test of the system with transmission to channel noise, the correlations of adjacent pixels in the encipher image, wider gray scale image database and its behavior in real time.

From an engineer's perspective, chaos-based image encryption technology is very promising for real-time secure image and video communications in biomedical, military and commercial applications.

REFERENCES

[1] G. Jakimoski and L. Kocarev, "Chaos and Cryptography: Block Encryption Ciphers Based on Chaotic Maps". IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications. Vol. 48, No. 2, 2001, pp. 163-169.

[2] C.C. Chang, M.S. Hwang and T.S. Chen, "A New Encryption Algorithm for Image Cryptosystems," Journal System Software, Vol. 58, 2001, pp. 83-91.

[3] H. Cheng and X. B. Li, "Partial Encryption of Compressed Images and Videos," IEEE Transactions Signal and Process, Vol. 48, No. 8, 2000, pp. 2439-2451.

[4] N. Bourbakis and C. Alexopoulos, "Picture Data Encryption Using SCAN Patterns," Pattern Recognition, Vol. 25, No. 6, 2007, pp. 567-581.

[5] H. H. Nien, C. K. Huang, S. K. Changchien, H. W. Shieh, T. Chen and Y. Y. Tuan, "Digital C and Decoding Using a Novel Chaotic Random Generator," Chaos Solitons and Fractals, Vol. 32, No. 3, 2005, pp. 1070-1080.

[6] Q. Alsafasfeh and A. Alshabatat, "Image Encryption Based on Synchronized Communication Chaotic Circuit," Journal of Applied Sciences Research, Vol. 7, No. 4, 2011, pp. 392-399.

[7] Q. Alsafasfeh and A. Arfoa, "Image Encryption Based on the General Approach for Multiple Chaotic Systems," Journal of Signal and Information Processing, Vol. 2, 2011, pp. 238-244.

[8] H. Gao, Y. Zhang, S. Liang and D. Li, "A New Chaotic Algorithm for Image Encryption," Chaos, Solitons and Fractals, Vol. 29, No. 2, 2006, pp. 393-399.

[9] C. Fu, Z. Zhang and Y. Cao, "An Improved Image Encryption Algorithm Based on Chaotic Maps," Third International Conference on Natural Computation, Vol. 3, Washington, 2007, pp. 24-27.

[10] L. Zhang, X. Liao and X. Wang, "An Image Encryption Approach Based on Chaotic Maps," Chaos, Solitons and Fractals, Vol. 24, No. 3, 2005, pp. 759-765.

[11] S. Bu and B. Wang, "Improving the Security Encryption by Using a Simple Modulating Method," Chaos, Solitons and Fractals, Vol. 19, No. 4, 2003, pp. 919-924.

[12] L. Shubo, S. Jing and X. Zhengquan, "An Improved Image Encryption Algorithm based on chaotic system", Journal of Computers, vol. 4, No. 11, November 2009, pp. 1091-1100.

[13] L. Wang, Q. Ye, Y. Xiao, Y. Zou and B. Zhang, "An Image Encryption Scheme Based on Cross Chaotic Map," Congress on Image and Signal Processing, Sanya, 27-30 May 2008, pp. 26-27.

[14] I. A. Ismail, M. Amin, H. Diab, "An Efficient Image Encryption Scheme Based on Chaotic Logistic Maps," International Journal of Soft Compution, Vol. 2, 2007, pp. 285-229.

[15] M. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients," IEEE Transactions on Signal Processing, Vol. 41, No. 12, 1993, pp. 3445-3462.

[16] T. Brahimi, A. Melit and F. Khelifi, "Fast Encryption Methods for Audiovisual Data Confidentiality," Digital Signal Processing, Vol. 19, 2009, pp. 220-228.

[17] R. Lin, Y. Mao and Z. Wang, "Chaotic Secure Image Coding Based on SPIHT," Third International Conference on Communications and Networking, China, 2008.

[18] M. Lahdir, A. Nait-Ali and S. Ameur, "Fast Encoding-Decoding of 3D Hyperspectral Images Using a Non-Supervised Multimodal Compression Scheme," Journal of Signal and Information Processing, Vol. 2, No. 4, 2011, pp. 316-321.

[19] S. Lian, J. Sun and Z. Wang, "A Secure 3D-SPIHT Codec," European Signal Processing Conference, Vi enna, Austria, 2004.

[20] A. S. Dmitriev, G.A Kassian and A.D Khilinsky, "Chaotic Synchronization of Henon Mappings: The Information Approach," Technical Physics Letters, Vol. 28, 2002.

[21] K. Vesely and J. Podolsky, "Chaos in a Modified Henon-Heiles System Describing Geodesics in Gravitational Waves," Technical Physics Letters A, Vol, pp. 271, 2000, pp.368-371.

[22] G. M. Maggio and O.D. Feo, "Nonlinear Analysis of the Colpitts Oscillator and Application to Design," IEEE Transactions on Circuits and Systems: Fundamantal Theory and Applications, Vol. 49, 1999.

[23] G. M. Maggio and M.P. Kennedy, "Experimental Manifestations of Chaos in the Colpitts Oscillator," in: Proc ICECS, Seville, Spain, 1997, pp. 194-204.

[24] I. Belmouhoub, M. Djemaï and J.P. Barbot, "Observability Quadtatic Normal Form for Discrete-Time Systems", IEEE Transactions on Automatic Control, Vol. 50, 2005.

[25] M. Djemaï, J.P. Barbot and I. Belmouhoub, "Discrete-Time Normal Form for Left Invertibility Problem", European Journal of Control, Vol. 15, pp. 194-204, 2009.

[26] H. Hamiche, M. Ghanes, J.P. Barbot and S. Djennoune, "Secure Digital Communication based on Hybrid Dynamical Systems," Communication Systems, Networks and Digital Processing, CSNDSP'10, Newcastle, U.K, 2010.

[27] F. Anstett, G. Millerioux and G. Bloch, "Chaotic Cryptosystems: Cryptanalysis and Identifiability", IEEE Transactions on Circuits and Systems: Fundamental Theory and Applications, Vol.53, 2006.

[28] H. Dimassi, A. Lori'a and S. Belghith, "A new secured scheme based on chaotic synchronization via smooth adaptive unknown-input observer," Communications in Nonlinear Science and Numerical Simulations, Vol. 17, No. 9, 2012, pp. 3727-3739.

AUTHORS PROFILE

Hamid Hamiche was born in Algeria in 1974. He received his Magister degree in Electronic from the Mouloud MAMMERI University of Tizi-Ouzou (Algeria) in 2002 and Ph.D. degree in Automatic Control from the Mouloud MAMMERI University of Tizi-Ouzou and National School of Electronics and Applications of Cergy-Pontoise (France) in 2011. His research activities deal with sliding-mode control, observation of chaotic systems and synchronization of chaotic systems. His main application domain is cryptography.

Mourad Lahdir was born in Algeria in 1969. He received his Magister degree in Electronic from the Mouloud MAMMERI University of Tizi-Ouzou (Algeria) in 1999 and Ph.D. degree in Electronics Remote Sensing from the Mouloud MAMMERI University of Tizi-Ouzou in 2007. His research activities are image processing, Meteosat and hyperspectral image compression, wavelet and fractal image application, progressive data transmission and watermarking.

Mohammed Tahanout was born in Algeria in 1975. He received his engineering degree in 1998 in Electronic telecommunications from the Mouloud MAMMERI University of Tizi-Ouzou (Algeria) and his Magister degree in Electronic radar system in 2003 from USTHB university of Algiers (Algeria). His research activities are radar image processing, telecommunications, radar system and signal processing.

Saïd Djennoune was born in Algeria in 1956. He received his Magister degree in Electronic from the High Commission for Research of Algiers (Algeria) and Ph.D. degree in Automatic Control from the Mouloud MAMMERI University of Tizi-Ouzou (Algeria). Since 2005, he is a Professor. His research activities deal with fractional derivative, synchronization of chaotic systems, control and observation of electric machines, which are applied to industrial problems.

Important Features Detection in Continuous Data

Piotr Fulmański, Alicja Miniak-Górecka
Faculty of Mathematics and Computer Science, Department of Mathematical Analysis and Control Theory
University of Łódź
Łódź, Poland

Abstract—**In this paper, a method for calculating the importance factor of continuous features from a given set of patterns is presented. A real problem in many practical cases, like medical data, is to find which parts of patterns are crucial for correct classification. This leads to the need of preprocessing all data, which has influence on both time and accuracy of applied methods (when unimportant data hide those which are important). There are some methods that allow selection of important features for binary and sometimes discrete data or, after some preprocessing, continuous data. Very often however, such conversion is burdened with the risk of losing important data, which is a result of lack of knowledge of optimal discretization consequence. Proposed method allows to avoid that problem, because it is based on original, non-transformed continuous data. Two factors - concentration and diversity - are defined and are used to calculate the importance factor for each feature and pattern. Based on those factors e.g. unimportant features can be identified to decrease dimension of input data or "bad" patterns can be detected to improve classification. An example how proposed method can be used to improve decision tree is given as well.**

Keywords-important features extraction; continuous data analysis; decision tree.

I. Introduction

In this paper, the following problem of the data processing and analysing is presented. Let L be a given learning set defined as:

$$L = \{l_1 = (p_1 = (c_1^1, \dots, c_m^1), t_1), \\ \dots, \\ l_n = (p_n = (c_1^n, \dots, c_m^n), t_n)\}. \tag{1}$$

L is a set of pairs $l_1, \dots, l_n$, where the first element (called: input signal) is an m-components vector of features (p_i, $i=1,\dots n$), while the second is a value which belongs to a given, finite set T. Notation c_j^i denotes j-th feature from i-th pattern. T is a set of correct (expected) output signals (also: responses, targets or classes). It can consist of numbers, but also of logic values: *yes*, *no*, *unknown* or linguistic: *brake*, *move slowly*, *move*, *accelerate*, *stop*. Features $c_1^i, \dots, c_m^i$, $i=1,\dots,n$ are independent of each other (i.e. set of values for feature c_p^i does not depend on set of values for feature c_q^i, $p \neq q$), can be both discrete and continuous.

Presented problem is solved when for each $t \in T$ there is known such a set of features, which is sufficient to unambiguous identification (classification) of all of the learning data for which t is an expected class. As an example, consider set L defined in table I. All patterns are divided into five different classes: A, B,…, E. Features which, according to our assumption, should characterize each class are embolden. Assumptions for set L were as follow.

- Class A should be recognized based on fact that feature 1 takes values from interval 10-30, whereas the rest of features should not have any regularity.
- Class B should be recognized based on fact that feature 1 takes values from interval 10-30 and features 2 and 3 take values from interval 50-65, whereas the rest of features should not have any regularity.
- Class C should be recognized based on fact that feature 2 takes values from interval 90-110, feature 3 takes values from interval 60-75, feature 4 takes values from interval 25-55, whereas the rest of features should not have any regularity.
- Class D should be recognized based on fact that feature 4 takes values from interval 0-25, whereas the rest of features should not have any regularity.
- Class E should be recognized based on fact that feature 1 takes values from interval 50-70, whereas the rest of features should not have any regularity.

According to the above assumptions a few randomly generated sets were created – the set L is one of them. In all cases results were similar.

TABLE I. EXAMPLE OF LEARNING SET *L*

Pattern	**Feature**					**Class**
	1	***2***	***3***	***4***	***5***	
p_1	**10**	65	50	50	50	A
p_2	**20**	70	60	25	70	A
p_3	**25**	80	100	95	130	A
p_4	**29**	100	90	100	105	A
p_5	**15**	110	**50**	**50**	80	B
p_6	**25**	90	**55**	**75**	55	B
p_7	**29**	60	**60**	**60**	60	B
p_8	**31**	75	**63**	**65**	150	B
p_9	5	**90**	**60**	**25**	110	C
p_{10}	30	**105**	**70**	**30**	145	C

Pattern	Feature					Class
	1	*2*	*3*	*4*	*5*	
p_{11}	15	**100**	**65**	**50**	60	C
p_{12}	58	**95**	**57**	**52**	45	C
p_{13}	95	**87**	**72**	**48**	50	C
p_{14}	100	**110**	**60**	**27**	90	C
p_{15}	40	70	60	**0**	60	D
p_{16}	70	80	55	**10**	70	D
p_{17}	80	100	95	**25**	120	D
p_{18}	**50**	110	95	10	70	E
p_{19}	**60**	60	80	60	80	E
p_{20}	**70**	75	50	110	110	E

II. Decision Tree and Continuous Data

From the previous section it can be seen, that the goal is to create a model that predicts the value of a target variable based on several input variables (features). As a predictive model a decision tree which maps observations about an item to conclude the target value of an item can be used. An interior node corresponds to one of the input variables; each of these nodes has a number of children nodes equal to the number of the possible values of that input variable. Each leaf node represents a possible outcome depending on the values of the input variables represented by the path from the root node to the leaf node. It is essential that a tree can be "learned" by splitting the source set into subsets based on an **attribute value test**. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion ends when the subset at a node has the same value of the target variable, or when further splitting no longer adds a value to the predictions.

In pseudocode, the general algorithm for building decision trees is [1]:

1. Check for base cases.
2. For each attribute *a* find the normalized information gain from splitting on *a*.
3. Let a_{best} be the attribute with the highest normalized information gain.
4. Create a decision node that splits on a_{best}.
5. Recur on the sublists obtained by splitting on a_{best}, and add those nodes as children of node.

In presented algorithm the most important are steps 2 and 3: selection a_{best} attribute. Selection of that attribute should be based on some factor describing its importance regarding data that are not classified yet. Term *importance* in this case is understood as an ability to create (based on that attribute) correct pattern classification -- the more patterns are classified correctly, the better (the more important) the attribute is. While for discrete data methods for attribute importance factor calculating were developed (see for example [2], where method for binary patterns recognition is described or C4.5 algorithm), the lack of such methods can be observed for continuous data. As an example of this problem consider one of the widely used free data mining tool i.e. C4.5 algorithm developed by Ross Quinlan [3] used to generate a decision tree and implemented in SIPINA Data Mining Software [4].

C4.5 is an extension of Quinlan's earlier ID3 algorithm and is followed in turn by See5/C5.0[1][5]. C4.5 made a number of improvements to ID3 -- one is important from our point of view: the ability to handle both continuous and discrete attributes. Unfortunately in order to handle continuous attributes, C4.5 creates a threshold and then splits the list into two: those which attribute value is above the threshold and those that are less than or equal to it [6]. As a result, continuous data are subject to some kind of discretization. This process can be performed before the main algorithm or as a one of auxiliary sub-steps of it. Anyway, continuous data are de facto treated as discrete. In many cases, discretization results in loss of information. In this paper, method for calculating importance factor of continuous features from given patterns set, without discretization necessity, is presented.

III. Measure of Importance of Features

While searching for important features that distinguish a given class among other classes, for each feature the following factors should be determined:

- if a feature is a distinctive feature within a given class (so-called importance factor for all patterns within a given class) -- for example, for all patterns this feature has the same value;
- if a feature is a distinctive feature for a given class within all classes (so-called importance factor for a given class within all classes) -- for example, for all patterns which are not from a given class this feature takes value from interval 0-10, while for patterns from a given class this feature takes value 15.

In a given examplary set of patterns *L* (table I) one can notice that feature 4 is the most important (the most distinctive) feature for class D within this class (the smallest diversity can be observed for it). Feature 4 is an example of second factor as the most important (the most distinctive) feature for class D within all classes, because for none of the other classes values of this feature belong to interval 0-25[2].

A. Importance factor for all patterns within a given class

For each feature, the smaller the changeability of its values within a given class is, the more important this feature is. In other words, concentration of this feature is higher. **Concentration factor** of feature *a* in class *b* is defined as:

[1] C5.0/See5 is a commercial and closed-source product. C5.0 offers a number of improvements on C4.5 like speed (C5.0 is several orders of magnitude faster than C4.5), more memory usage efficient or smaller decision trees (C5.0 gets similar results to C4.5 with considerably smaller decision trees).

[2] Values from this interval that can be observed for feature 4 in other classes e.g. pattern 2 (class A) with value 25 or pattern 18 (class E) with value 10 simulate anomalies in the data and were added intentionally.

$$cf_a^b = \int_{-\infty}^{+\infty} exp\left(\frac{-\left(x-\mu_a^b\right)^2}{2\sigma_a^{b^2}}\right) dx, \qquad (2)$$

where μ_a^b is a mean (expected value) and σ_a^b is a standard deviation of all values for feature a in class b. The smaller the concentration factor is, the closer the values of a considered feature within a given class are. It can be interpreted in the following way: if all values of a considered feature within a given class are (almost) identical, it can be stated that this feature (its values) is being characteristic for all patterns within a given class.

For example a characteristic feature of all tanks is to have tracks (but not all tracked vehicle are tanks). Examining concentration factors for patterns from set *L* (see table II), one can notice that for each class the smallest value of this factor is located in one of the features, which were assumed to be characteristic. It is worth highlighting that the set *L* is not "perfect" -- as some patterns are not necessarily fulfilling all assumptions for a corresponding class to which these patterns belong in a way that a wrong classification would be excluded. For example, pattern 16 (from class D) could be assigned to class E.

B. *Importance factor for a given class within all classes*

A feature is considered to be the more diversified, the greater changeability of its values within all classes is observed. **Diversity factor** of feature *a* within all classes is defined as:

$$cf_a = \int_{-\infty}^{+\infty} exp\left(\frac{-(x-\mu_a)^2}{2\sigma_a{}^2}\right) dx, \qquad (3)$$

where μ_a is a mean (expected value) and σ_a is a standard deviation of all values for feature *a* within all classes. Diversity factor is a little bit more difficult to describe than concentration factor. It has much more sense when considered jointly with the concentration factor (see next subsection). For now, we can say that a small value of this factor means that many patterns from different classes take similar values.

TABLE II. CONCENTRATION FACTORS FOR PATTERNS FROM SET *L*. THE SMALLEST VALUE FOR EACH CLASS IS UNDERLINED.

Feature					Class
1	***2***	***3***	***4***	***5***	
17.81	33.60	51.67	78.51	77.44	A
15.45	46.36	**12.40**	**22.59**	95.19	B
92.87	**20.23**	**13.80**	**28.78**	89.69	C
42.60	31.26	44.60	**25.75**	65.79	D
20.46	52.51	46.89	102.33	42.60	E

TABLE III. DIVERSITY FACTORS FOR PATTERNS FROM SET *L*.

Feature				
1	***2***	***3***	***4***	***5***
69.40	41.19	39.26	74.42	79.84

C. *Discriminants*

A discriminant describes how important a given feature of the considered pattern is for its correct classification. Discriminants are calculated for all features of all patterns with the following formula:

$$d_a^{b,c} = \frac{df_a}{cf_a^b} exp\left(\frac{-(x-\mu_a^b)^2}{2\sigma_a^{b^2}}\right), \qquad (4)$$

where *x* is a value of feature *a* from pattern *c* and class *b*. In formula (4) two component can be distinguished.

- The first component is a quotient which is calculated for each feature as a diversity factor for a given class (and feature) within all classes over concentration factor for all patterns within a given class (and feature). Value of this quotient close to 1 means that the feature which is being under consideration cannot be treated as a characteristic feature (discriminant) for the class. The most desirable is a "big" value of this component, which is obtained when values of a given feature in a selected class compared to values of this feature in other classes are evidently concentrated, that is when a feature is perfect to act as a characteristic (discriminant) of the class. This component is being calculated for every feature in all classes (see table IV).
- The second component, *exp()*, serves to eliminate data which are (very) different from the average value for a given class, that is data which could be an effect of measuring errors or some kind of an anomaly which should be considered individually. A value of this component close to 0 means that the feature in a considered pattern is greatly deviated from the average value for an appropriate class. On the other hand, when the value of this component is close to 1 it means that the feature in a considered pattern has a typical value for an appropriate class. In other words, **second component describes the grade of membership of a feature in a given pattern to the usual values of this feature in patterns from an appropriate class.** Averaging all grades of membership of features of a pattern, the *grade of membership of a pattern to a class* is obtained, which is denoted as $\mu^{b,c}$, where *c* - patterns, *b* -- class. Knowledge of the grades of membership of patterns is useful for "bad" patterns identification. Values of this component and the grades of membership are given in table V.

Taking into consideration the total effect of described elements, one can state that values calculated with formula (4) lower or equal to 1, shows features which should not be considered.

If this value is greater than 1 (the more, the better) then the considered feature is important. Final values of the discriminants for set *L* are presented in table VI.

The greatest value for each pattern is underlined. It can be noticed, that in all cases discriminants reach the greatest value for a feature which, according to initial assumptions, should be characteristic for a given class.

TABLE IV. VALUE OF QUOTIENT $\frac{df_a}{cf_b^a}$ FOR DATA FROM TABLE II AND III.

Feature					Class
1	*2*	*3*	*4*	*5*	
3.89	1.12	0.75	0.94	1.03	A
4.49	0.88	**3.16**	**3.29**	0.83	B
0.74	**2.03**	**2.84**	**2.58**	0.89	C
1.62	1.31	0.88	**2.89**	1.21	D
3.39	0.78	0.83	0.72	1.87	E

TABLE V. THE GRADES OF MEMBERSHIP OF FEATURES AND PATTERNS.

Feature					$\mu^{b,c}$	Class
1	*2*	*3*	*4*	*5*		
0.3	0.59	0.47	0.85	0.45	0.532	A
0.99	0.8	0.76	0.39	0.83	0.754	A
0.85	0.99	0.47	0.68	0.41	0.68	A
0.53	0.28	0.76	0.58	0.87	0.604	A
0.26	0.36	0.36	0.38	0.71	0.79	B
1.0	0.94	0.92	0.38	0.71	0.79	B
0.81	0.43	0.83	0.96	0.78	0.762	B
0.62	0.89	0.47	0.96	0.24	0.636	B
0.47	0.62	0.76	0.49	0.75	0.618	C
0.85	0.67	0.55	0.75	0.22	0.608	C
0.63	0.96	0.98	0.61	0.8	0.796	C
0.97	0.94	0.44	0.5	0.56	0.682	C
0.48	0.4	0.34	0.71	0.64	0.514	C
0.4	0.32	0.76	0.59	0.98	0.61	C
0.38	0.56	0.85	0.52	0.67	0.596	D
0.92	0.96	0.7	0.98	0.87	0.886	D
0.61	0.4	0.37	0.43	0.37	0.436	D
0.47	0.4	0.56	0.47	0.61	0.502	E
1.0	0.58	0.96	1.0	0.92	0.892	E
0.47	0.95	0.4	0.47	0.38	0.534	E

In case of classes A, D and E one feature was selected explicitly: first, fourth and first respectively. Explicitness is not observed in case of class B and C. For class B first feature (once) and fourth feature(twice) was detected as the most characteristic. For class C: third (three times), fourth (twice) and second (once). Those inconsistencies signal the need for usage of more features in case of some classes. In table VII the second highest discriminants relative to the values of discriminant for class B and C are shown (these values are underlined; for clarity, the highest value for each pattern is removed).

TABLE VI. DISCRIMINANTS FOR PATTERNS FROM SET *L*. THE HIGHEST VALUE FOR EACH PATTERN IS UNDERLINED.

Pattern	Feature					Class
	1	*2*	*3*	*4*	*5*	
p_1	**1.17**	0.72	0.36	0.81	0.46	A
p_2	**3.85**	0.99	0.58	0.37	0.85	A
p_3	**3.32**	1.22	0.36	0.64	0.42	A
p_4	**2.06**	0.34	0.58	0.55	0.89	A
p_5	**1.2**	0.32	**1.16**	**1.25**	0.82	B
p_6	**4.49**	0.83	**2.91**	**1.25**	0.59	B
p_7	**3.63**	0.38	**2.63**	**3.16**	0.66	B
p_8	**2.79**	0.79	**1.51**	**3.16**	0.2	B
p_9	0.35	**1.27**	**2.18**	**1.27**	0.67	C
p_{10}	0.64	**1.37**	**1.57**	**1.94**	0.2	C
p_{11}	0.47	**1.96**	**2.79**	**1.58**	0.71	C
p_{12}	0.73	**1.91**	**1.26**	**1.31**	0.5	C
p_{13}	0.36	**0.82**	**0.99**	**1.85**	0.57	C
p_{14}	0.3	**0.65**	**2.18**	**1.54**	0.87	C
p_{15}	0.63	0.74	0.75	**1.51**	0.81	D
p_{16}	1.5	1.27	0.61	**2.85**	1.06	D
p_{17}	1.0	0.53	0.32	**1.24**	0.45	D
p_{18}	**1.6**	0.31	0.47	0.34	1.15	E
p_{19}	**3.39**	0.45	0.8	0.72	1.73	E
p_{20}	**1.6**	0.74	0.34	0.34	0.73	E

TABLE VII. THE SECOND HIGHEST DISCRIMINANTS (UNDERLINED) FOR PATTERNS FROM SET *L*. FOR CLARITY, THE HIGHEST VALUE FOR EACH PATTERN IS REMOVED.

Pattern	Feature					Class
	1	*2*	*3*	*4*	*5*	
p_5	**1.2**	0.32	**1.16**		0.82	B
p_6		0.83	**2.91**	**1.25**	0.59	B
p_7		0.38	**2.63**	**3.16**	0.66	B
p_8	**2.79**	0.79	**1.51**		0.2	B
p_9	0.35	**1.27**		**1.27**	0.67	C
p_{10}	0.64	**1.37**	**1.57**		0.2	C
p_{11}	0.47	**1.96**		**1.58**	0.71	C
p_{12}	0.73		**1.26**	**1.31**	0.5	C
p_{13}	0.36	**0.82**	**0.99**		0.57	C
p_{14}	0.3	**0.65**		**1.54**	0.87	C

Taking into consideration those two features (one and four), the correct classification for class B should be possible.

Class C can be still a source of problems, because different pairs of features were selected: feature two and three (twice), feature three and four (three times) and finally feature two and four (once). Nothing prevents the next feature (the third highest discriminant) from being considered.

As a result, for class C features two, three and four will be selected[3]. Features one, three and four are selected if for class B three features are also considered.

Notice that based on values from table IV importance of features can also be estimated. However, information about sets of features, as it was described above, cannot be determined. Therefore data from table IV can be treated as a rough selection of important features, while richer information is contained in discriminants calculated with formula (4) (see table VI).

IV. Usage Example

In this section an example how the proposed method can be used to improve a decision tree is given. A decision tree generated in SIPINA [4] tool (C4.5 algorithm was selected) for learning set *L* is presented on Fig. 1.

It can be noticed, that feature five was not considered in any nodes, which could be predicted by analyzing table VI. Feature one is the first feature that splits the data set. Afterwards, feature three and four are considered. This is also reflected in table VI.

Knowledge of the grade of membership $\mu^{b,c}$ of pattern *c* to class *b* (how representative the selected pattern is for that class) allows one to modify learning set in such a way that smaller classification error will be achieved. For the considered learning set *L*, the smallest grades of membership are for patterns p_5 (0.468), p_{17} (0.436) and p_{18} (0.502).

One can notice that the decision tree from Fig. 1 does not make correct classification for all data. Data which are classified ambiguously are: (p_2, p_9), (p_1, p_5) and (p_{15}, p_{16}, p_{17}, p_{18}).

However, there is a correct classification for every case for reduced learning set *L* (patterns p_5, p_{17} and p_{18} were removed; see Fig. 2). Of course reduction of the data learning set may not have a permanent and strict character - it can be treated as a selection of potentially problematic patterns which should be treated separately.

V. Conclusions and Plans

All sets which were used during the tests (presented learning set *L* is one of them) are characterized by

- randomly generated set of features according to some assumptions which was described in section 1;
- existence of contradictory data - pattern 16 could just as well belong to class E and pattern 18 to class D.

In all cases the presented method for important features detection in continuous data works well. All features which, according to our assumptions, should be important were identified as such. The grade of membership usage allows more effective utilization of a data learning set through isolation of potentially problematic patterns (which could e.g. have negative influence during classification process). Notice, that global knowledge of important features gives new abilities. Instead of splitting data based on one feature (like in decision tree), a set of them (the most important) can be used to improve the decision process.

We want to stress, that in this paper an answer for a question: *which features are essential for correct pattern classification of a given class* is given. Proposed method is not a complete tool for data classification - it can be considered as an element of such system. This will be our next research problem - how to use information about important features to build classification system for a really problematic data, like medical data, which in many cases are incomplete or contradictory. Additionally, a new problem that we want to investigate arose: how to treat incomplete patterns.

References

[1] S. B. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques", Informatica, vol. 31, 2007, pp. 249-268.

[2] A. Horzyk, "Fast Automatic Configuration of Artificial Neural Networks Used for Binary Patterns Recognition", Biocybernetics and Biomedical Engineering, vol. 21, 2001.

[3] J. R. Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, 1993.

[4] R. Rakotomalala, SIPINA, http://eric.univ-lyon2.fr/~ricco/sipina, 2010.

[5] RULEQUEST RESEARCH, "Is See5/C5.0 Better Than C4.5?, http://www.rulequest.com/see5-comparison.html, 2009.

[6] J. R. Quinlan, "Improved use of continuous attributes in c4.5.", Journal of Artificial Intelligence Research, 1996, pp.77-90.

[3] Feature five selected by pattern p_{14} is omitted - we treat it as an anomaly.

)

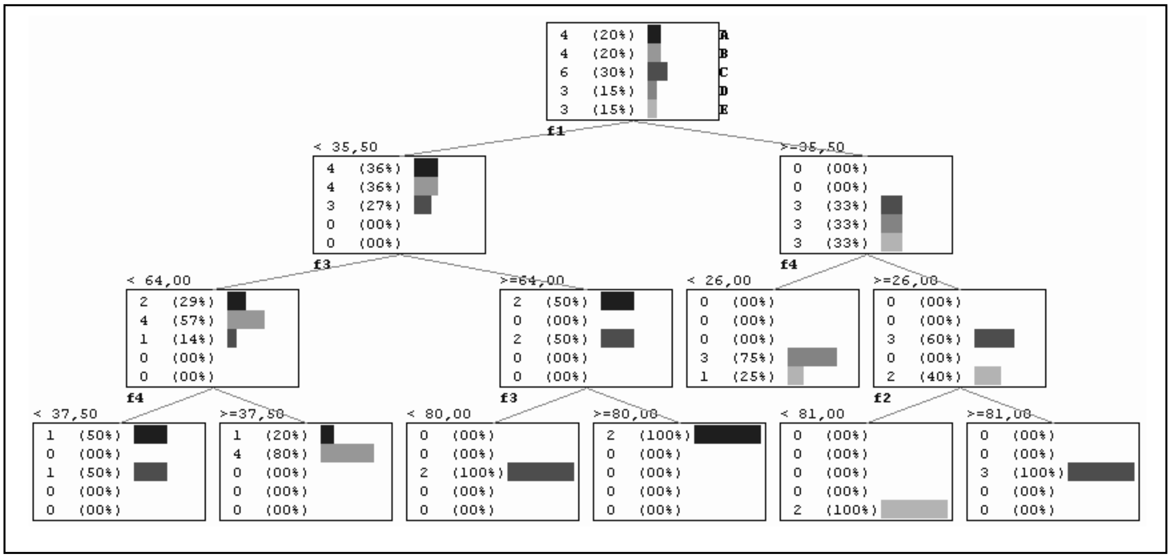
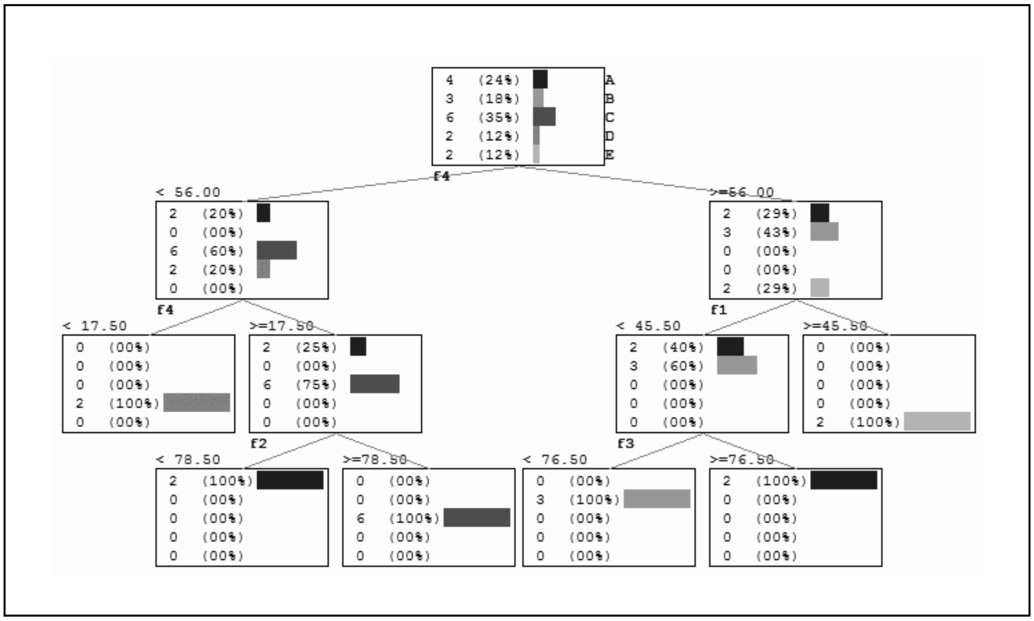


Figure 1. Decision tree generated in SIPINA tool (with C4.5 algorithm) for the learning set *L*.



)

Figure 2. Decision tree generated in SIPINA tool (with C4.5 algorithm) for a reduced learning set *L*.