IJACSA

WHERE WISDOM SHARES

International Journal of Advanced Computer Science and Applications

Volume 4 Issue 1

January 2013

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…

It is our pleasure to present to you the January 2013 Issue of International Journal of Advanced Computer Science and Applications.

Today, it is incredible to consider that in 1969 men landed on the moon using a computer with a 32-kilobyte memory that was only programmable by the use of punch cards. In 1973, Astronaut Alan Shepherd participated in the first computer "hack" while orbiting the moon in his landing vehicle, as two programmers back on Earth attempted to "hack" into the duplicate computer, to find a way for Shepherd to convince his computer that a catastrophe requiring a mission abort was not happening; the successful hack took 45 minutes to accomplish, and Shepherd went on to hit his golf ball on the moon. Today, the average computer sitting on the desk of a suburban home office has more computing power than the entire U.S. space program that put humans on another world!!

Computer science has affected the human condition in many radical ways. Throughout its history, its developers have striven to make calculation and computation easier, as well as to offer new means by which the other sciences can be advanced. Modern massively-paralleled super-computers help scientists with previously unfeasible problems such as fluid dynamics, complex function convergence, finite element analysis and real-time weather dynamics.

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.

Lastly, we would like to express our gratitude to all authors, whose research results have been published in our journal, as well as our referees for their in-depth evaluations.

We hope that materials contained in this volume will satisfy your expectations and entice you to submit your own contributions in upcoming issues of IJACSA

Thank you for Sharing Wisdom!

Managing Editor
IJACSA
Volume 4 Issue 1 January 2013
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

Reviewer Board Members

CONTENTS

Classifying Personalization Constraints in Digital Business Environments through Case Study Research

Michael J. Harnisch
Institute of Information Science and Information Systems
University of Graz / evolaris next level GmbH
Graz, Austria

***Abstract*— To aid professionals in the early assessment of possible risks related to personalization activities in marketing as well as to give academics a starting point to discover not only the opportunities but also the risks of personalization, a 'Classification Scheme of Personalization Constraints' is established after the analysis of 24 case studies. The classification scheme includes three dimensions: origin (internal, external), subject (technological, organizational) and time (data collection, matchmaking, delivery) and describes the different obstacles with which companies are confronted when implementing personalization activities. Additionally four 'Standard Types of Personalization Environments' are developed. They describe a set of business environments which inherit different internal and external risks related to personalization activities in marketing. The standard types are termed Flow, Performance, Dependence and Risk.**

Keywords—personalization; case study research; corporate communication; personalization constraint

I. INTRODUCTION

Developments in communication technology changed the communication patterns of customers and corporations. The traditional transfer of messages from person to person was redeemed by mass media communication which is nowadays more and more replaced by individualized respectively specified one-to-one-communication [1]. Additionally, an enormous amount of data and information is publicly available through the Internet, which is accessible through various stationary and mobile devices [2]. These developments equip customers with the ability to retrieve more information about a product or service much faster. But there are also downturns of these advances. Customers struggle with identifying relevant information which fit their personal preferences and needs [3]. Therefore companies need to implement certain actions in their corporate communication processes so that they communicate relevant marketing messages to their customers in an effective way. These actions could imply personalization activities, which include for example an individual adaption of customer touch points to the actual context or preferences of the customer, the provision of specific individualized messages or the collaborative-aided recommendation of similar products or services which the customer is looking for. By applying personalization activities on customer-related business processes, corporations are following a differentiation strategy [4-6]. This could yield additional value to the business by making communications more effective and as a result, raise for example conversion rates or buying intentions of the customers. Though, the application of personalization activities is subject to various limitations. For example, even if specialized tracking systems are employed, corporations are not always able to identify the context or the preferences of their customers. Businesses could also be confronted with legal constraints of implementing personalization activities (e.g. Opt-In requirements) or they suffer from elementary problems like privacy concerns or a lack of customer trust in personalization activities [7]. Research shows, that these constraints are able to significantly change the successful outcome of personalization activities.

Personalization is a very broad issue in research and ranges from computer sciences to social sciences. Although the influences of personalization constraints on the successful implementation of personalization activities and furthermore on the profit of businesses are apparent, only limited research has been published on the general theory of personalization constraints. A feasible explanation of this fact is the widespread possibilities of detailed research on personalization issues. Nevertheless, a classification scheme of personalization constraints is able to provide a first general overview of the various difficulties a company could come across while implementing personalization activities in corporate communication processes. The classification scheme is a relevant contribution for practitioners but also a starting point for further academic research which could provide a better understanding of the different issues during the implementation of personalization activities. Before the 'Classification Scheme of Personalization Constraints' and four 'Standard Types of Personalization Environments' are presented as the final result of this contribution, the approach which led to this results is depicted. At the beginning, a literature review has been conducted. A subsequently applied scientometric approach enhanced the findings of the literature review and served as a supplier of additional insights for the development of the morphological box of personalization. The box was used as an instrument to cover all relevant aspects within the performed case study analysis (minimum and maximum case deviation). The cases were analyzed to generally prove the developed classification scheme and the standard types of personalization as well as to fill possible remaining theoretical gaps within this model.

The structure of the paper starts with a literature review on personalization in general and the definition of 'constraints of personalization activities' in section 2. It is followed by a description of the used methods in section 3 before the results of the research are presented and discussed subsequently in section 4. Section 4 is divided into an overview of the results of the scientometric approach (section 4.1), the morphological box and its description (section 4.2) and the discussion of the findings of the case study research which includes the general description of possible obstacles in personalization activities, an extract of analyzed cases as well as the final results the 'Classification Scheme of Personalization Activities' and the four 'Standard Types of Personalization Environments' (section 4.3). The paper is concluded with the limitations of the study, a summary of the basic findings and an outlook to possible further research.

II. Literature Review On Personalization

The literature review on personalization research shows, that it is a very broad topic which spans across different research fields [8-11]. Research questions are ranging from the effective application of recommender systems in computer sciences [12] to the perceived privacy threats in social sciences and business informatics [7]. Extensive research has been conceded to these problem sets, which resulted in very specialized research in every field. Nevertheless, based on these circumstances, personalization research largely fails to provide a commonly agreed general theory. Especially when it comes to the definition of 'personalization', different descriptions have emerged throughout the last years (see Table I).

TABLE I. Definitions Of Personalization

Source	Definition
[13]	"Personalization is a process of providing relevant content based on individual user preferences, and personalized web sites obtain preferences information implicitly by tracking customer purchase or usage habits."
[14]	"Personalization is a firm's decision on the marketing mix suitable for the individual that is based on previously collected customer data."
[15]	„Personalization is the adaption of products and services by the producer for the consumer using information that has been inferred from the consumer's behavior or transactions."

Only few contributions are focusing on a general viewpoint on personalization research [8-10] [15-16]. The main findings of their research are for example a distinction between personalization and customization [9], various classification schemes and frameworks of personalization activities [8-10] [17-18] as well as standard types of personalization [8]. A classification scheme which has to be emphasized (see Table II) divides personalization activities in three dimensions: object (what), target (to whom) and origin (who) [8].

The standard types of personalization include an architectural perspective, which aims on the functionality of the web environment; an instrumental perspective, which gives the emphasis on the efficiency and productivity; a relational perspective, that covers social interaction and privacy and a commercial perspective, which intends to increase sales and customer loyalty [8].

TABLE II. Classification Scheme Of Personalization [8]

	Implicit	**Explicit**
Content	Individuated Categorical	Individuated Categorical
User Interface	Individuated Categorical	Individuated Categorical
Channel / Information Access	Individuated Categorical	Individuated Categorical
Functionality	Individuated Categorical	Individuated Categorical

Businesses are well advised if they take the different perspectives of personalization activities into consideration. This is especially true if they apply personalization as differentiation strategy, which aims to be a competitive advantage over competitors [5]. The standard types of personalization indicate that there are various types of differentiation when applying personalization activities, ranging from dynamic pricing to an individualized system design. Applying personalization is especially popular in e-Commerce environments to attract and retain customers [4] [6]. The communication of individualized messages is seen as an important element of the online marketing mix [19].

But corporations need to handle certain obstacles and constraints when implementing personalization technologies and strategies into their processes. They can be grouped into technological and organizational constraints, in which various subcategories can be found, like judicial or informational obstacles. Although personalization research is a major source of solutions to overcome the mentioned difficulties, a general definition of constraints of personalization activities as well as a general classification of possible constraints is missing. Hence it is necessary to generally define constraints of personalization activities based on the findings about personalization in the literature.

Concluding the discussion, a personalization constraint in corporate communications is every internal or external interference of a business to provide differentiated communication of information based on individual, stated or implied preferences of the customer.

III. Methodolody

To define the foundations of personalization as a starting point of this study, an extensive literature review has been conducted. Additionally, a scientometric approach was applied to enhance the results and give further insights into popular research topics in personalization and subsequently aid in the development of the morphological box. Therefore, the scientific database 'Scopus' was used to extract the 200 most cited articles published from 2002-2011 (20 from each year) which are related to personalization. The database is part of the SciVerse Platform and provides access to the titles and abstracts of publications in high-ranked journals in the research field. Naturally, not all relevant journals are included in a single database. Nevertheless, due to the reason, that the detection of the most discussed issues in personalization activities of corporations is the aim of the approach, and the database gives an overview about personalization topics, the provided dataset seems to be sufficient. Subsequently, different parameters and characteristics of personalization were found by introducing a morphological box [20]. Based on these findings, a general model of personalization constraints, a 'Classification Scheme of Personalization Constraints' and four 'Standard Types of Personalization Environments' were developed which served as the underlying hypothesis of the study.

To prove the developed models or theory respectively, a multi case study analysis was conducted based on the methodological approach of [21] and [22]. The general aim of case study research is to 'expand and generalize' the developed theory by an analytic generalization [21]. This approach is especially advantageous when a complex and dynamic field is examined and if a theory is derived from the analyzed cases [22], which both seems true for the research area in focus. Similar approaches already led to significant developments in other scientific areas [23]. To cover all relevant aspects of personalization by the multi case study approach, a list of cases was set up which was balanced based on the different parameters of the morphological box (minimum and maximum case deviation).

Afterwards, the structure of the individual case reports was defined. They include a general description of the case, a classification of the case by means of the morphological box, a detailed list and description of the found personalization constraints as well as a classification of the case based on the developed theories. Successively, the relevant data from 24 cases was collected and analyzed by applying a web content analysis. Afterwards, the stated theory was adapted to describe the general findings based on the single reports. Subsequently, the single case study report structure was amplified to reflect the adapted theory and the analyzed cases were updated. Finally, a cross case analysis was performed and the additional insights were transformed on the model.

As a result of the gathered data sets several findings can be presented. First, a general differentiation of personalization constraints was developed. Subsequently, a 'Classification scheme of Personalization Constraints' for businesses was proved and finally four 'Standard Types of Personalization Environments' were proposed to enhance the existing personalization theory and aid businesses in successfully identifying possible threats and constraints when planning to apply personalization actions and thus avoid major mistakes in the process.

IV. RESULTS

A. Recognized Issues in Personalization Research

As a method to enhance the findings which were retrieved through the conducted literature review, a scientometric approach was used to identify the most relevant and recognized issues in personalization research during the last ten years.

First, the most cited publications of the last ten years were retrieved by using the scientific database 'Scopus'.

In this database, which is a source of leading peer-reviewed journals, the term 'personali?ation' was searched. The questionmark was used as a wildcard, which could return results for any applicable character. Due to that approach, the British as well as the US notations of personalization were recognized. The subject area was restricted to 'Computer Science', 'Engineering', 'Social Sciences', 'Mathematics', 'Business, Management and Accounting', 'Decision Sciences', 'Arts and Humanities' and 'Economics, Econometrics and Finance' due to the reason that especially medical articles are not part of the research focus and should be excluded. After retrieving all 5,316 articles in the database related to the relevant part of personalization research, for each of the last ten years (2002-2011), the 20 most cited journal articles were selected. The number of citations ranged from 4 to 288.

The database provided only abstracts for the selected articles, but this seems sufficient in the light of the research objective due to the reason, that the main findings and issues of an article are provided in the abstract. A content analysis was performed on the 200 most cited abstracts related to personalization. They were exported to the open source content analysis tool 'TagCrowd.com', which was used to highlight the 50 most used terms for each year as well as the 50 most used terms for the whole period. An overview of the outcome is provided in Table III and Table IV.

The results of the introduced scientometric approach on personalization research shows, that – naturally – the user is the focus point of personalization issues. But it is also found, that the retrieval of correct information and data in web environments through learning systems is a frequently addressed topic throughout the last ten years. Personalization research also covers issues related to services and the design of certain systems and products.

Finally it is found, that collaborative filtering approaches, social media contexts, data mining techniques and adaptive systems are used to identify and satisfy the preferences of customers. These results were incorporated into the theory design as well as into the development of the morphological box.

TABLE III. TOP THREE RECOGNIZED TERMS PER YEAR

2011	model (29)	personalization (28)	learning (25)
2010	user (53)	systems (32)	tags (27)
2009	user (68)	personalized (44)	information (25)
2008	user (51)	learning (28)	recommendation (26)
2007	user (56)	web (35)	personalization (31)
2006	user (49)	personalization (32)	web (28)
2005	information (39)	based (29)	personalization (28)
2004	user (54)	recommendation (27)	search (24)
2003	personalization (27)	web (25)	system (24)
2002	user (30)	personalization (27)	web (19)

TABLE IV. MOST RECOGNIZED TERMS 2002-2011

user (305)	personalization (245)	information (185)	web (169)	learning (161)
results (152)	systems (152)	based (129)	paper (114)	recommendation (113)
search (112)	data (111)	model (111)	research (107)	approach (104)
services (102)	study (95)	interests (89)	design (88)	algorithm (82)
present (82)	query (82)	different (81)	provide (77)	techniques (76)
used (73)	collaborative (72)	content (69)	applications (67)	knowledge (66)
preferences (66)	items (64)	analysis (62)	effective (61)	important (61)
social (61)	customers (60)	trust (60)	framework (59)	context (58)
mining (56)	proposed (53)	filtering (51)	online (51)	consider (50)
adaptive (49)	article (49)	people (49)	process (48)	improving (47)

B. Results And Description Of The Morphological Box

After defining the relevant and popular topics in personalization research, the 'Morphological Box of Personalization' was developed to enhance the output of the following case study analysis and aid the choice of observed cases (see Table V). The morphological box is recognized as a creativity method to cover all aspects of a defined issue [20]. At this stage of the study, is has been used to define all relevant parameters of personalization activities and their value to aid the subsequent case studies in depicting the possible obstacles.

By following the chosen classification scheme of personalization activities by [8], the first parameters of the 'Morphological Box of Personalization' were found. As one of the first decisions, the company has to choose which object should be personalized. It can include content, user interface, functionality and channel [8]. These defined parameter values can be enhanced by certain subcategories like for example the personalization of price [15] [24], which is attributable to content personalization.

TABLE V. MORPHOLOGICAL BOX OF PERSONALIZATION

parameter	**value**					
object	content	user interface	functionality	channel		
target	individual (1:1)		categorical (1:n)			
origin	user-driven (explicit)		company-driven (implicit)			
motive	cognitive	affective	social	self-expression		
strategy	individualization	utilization	segmentation	mediation		
focus	person-related		context-related			
media	print	radio	TV	Internet	mobile	other
aim	revenue	response rate	loyalty	satisfaction	differentiation	
cost	transaction cost	time	premium rate	data	other	
filtering	rule-based	content-based	collaborative filtering	hybrid	other	

Subsequently, the degree [25] or target [8] of personalization needs to be selected. It can either be a specialized personalization of the object for a single individual or a categorical personalization, which targets a classified group of persons. It has to be mentioned, that a more specified personalization is not always efficient [26].

Additionally, a differentiation between user-driven (explicit) and company-driven (implicit) personalization can be stated [8] [27]. The explicit origin of personalization, which builds on self-revealed information of the user includes the identification profile (e.g. name, contact data), the preference profile (self-revealed preferences), the socio-economic profile (e.g. age, gender), ratings of products, reviews or pages as well as relationships to other users/customers and given reviews and opinions. On the other side, the implicit personalization which is done automatically by an IT-agent or system of the company can include a transaction profile with a transaction log, an interaction profile (click-stream data) or external data like news or the weather report [25] [27-28].

After defining the basic foundations of each personalization activity, the company now has to choose, which motive the personalization activity underlies respectively which value it would like to create for the customer by personalizing its' offerings. In general it is possible to derive four fundamental needs or motives a customer follows when consuming media. Personalization activities should strive to serve the desired needs best. They split into cognitive motives, which include information about products or services, affective motives, which aim on the entertainment needs of the customer, social motives, which enable the customer to communicate with others and motives of self-expression, which assist the customer in constructing his personal self [29-30].

Companies need to visualize the strategy they would like to follow when implementing personalization. Personalization strategies can be divided into individualization, utilization, segmentation and mediation. Individualization strategies aim to provide a best suited and personalized design 'that incorporates the needs and requirements of users' [8] to enhance the quality and functionality. When utilization is chosen, efficiency is assured by using the right channel and media to deliver the information. The segmentation strategy aims on segmenting the relevant market and users into groups and provide them differentiated products, services or information. Finally the mediation strategy ensures the best possible linkage for social interaction between individuals and strives to enable them to expand their personal relationships [8].

Another differentiation in personalization activities is related to the focus, which can be person-related and is therefore in a close relationship to the collected user (transaction) data, or context-related, which aims on targeting the corporate information on the actual context of the customer like time of the day, current position, date or actual activities performed [25].

Personalization activities naturally need underlying media where they are performed on. They could include classic print media like newspapers or magazines, radio, TV, the Internet in general like websites or social media, mobile devices or others. For this research purpose, digital environments like the Internet or communication with mobile devices is relevant.

Another important issue is the aim of the personalization activity. They can be aimed on growing revenue in general or on improving the response rate of customers. But activities could also be directed on the customer loyalty or the customer satisfaction. Finally by doing a personalization activity, also differentiation of the offered product, service, information or the company itself could be an aim [11].

It is also important to depict the costs of the activity for the user. They can include simple transaction costs like e.g. for the access to the Internet. If data is provided by customers, they naturally need time to complete questionnaires or to state their preferences. The submission of additional data, which could be necessary for some personalization activities, implies data security costs. In some cases, customers are charged premium rates for personalization. Other costs are also conceivable [11].

Finally, there is also a difference in the filtering which is applied when personalization activities are performed. First, rule-based filtering for relevant information could be employed, which generates results based on static rules which could be related to all aspects of the user profile like historical transactions or demographic data or to a certain user choice [6]. Also content-based filtering [16] [25] could be applied, which lays the focus on the information which should be provided. Collaborative filtering methods take the preferences, transactions and ratings of other users for the provision of relevant information to the customer into consideration [15-16] [25]. Hybrid forms of filtering are also available, which combine different filtering approaches [16]. And finally, there are other approaches too, which include for example web usage mining methods [31].

C. Case Study Analysis

As a result of the case study analysis, different types of personalization constraints are introduced. In general, they can be grouped into internal and external constraints.

Internal constraints of personalization activities assemble all types of obstacles that are based on business-internal issues, which detain the corporation from successfully implementing a personalization activity. They can include technological issues like the inability to deploy a certain personalization system or organizational constraints like e.g. that the general characteristics of the sold product or service are not efficiently personalizeable.

External constraints of personalization activities on the other hand gather issues which are not lying in the inability of the company, but the customer or other external institutions or individuals. Examples are statutory rules about the usage of personal data for personalization activities or the attitude of a customer about the provision of individual preferences for a certain application.

There are several classes of personalization constraints that can be drawn. Inside these classes, the specific constraints are loosely related to the steps in the personalization process proposed by [11] and [16]. On the one hand there are technological constraints, which are divided into technological constraints of data collection, technological constraints of matchmaking and technological constraints of delivery. Technological constraints of data collection arise when a data collection system is not able to retrieve the data needed to provide a sufficient personalization base. When the matchmaking system is not able to identify the best information, products or services because it is not able match the offerings with the preferences of the customer, technological constraints of matchmaking occur. Finally, if it is technically not possible to deliver or present the results respectively the design which fits the needs and preferences of the customer, technological constraints of presentation happen.

On the other hand, organizational constraints of personalization activities arise. They include organizational constraints of data collection, organizational constraints of matchmaking as well as organizational constraints of delivery. Data collection could for example be limited by data privacy issues or by the circumstance that not all of the data which has been retrieved from the customer is useful for personalization activities. Matchmaking constraints include the inability to find a suitable product or service which fits the needs of the customer because for example the company does not offer a product variation which applies to the customers' preferences. Finally organizational constraints of delivery are found if for example, the timing of the result presentation does not meet the customers' ability to receive a marketing message.

The following list (see Table VI) includes examples of analyzed cases where businesses have successfully implemented personalization activities. The list contains all cases which are used in the 'Classification Scheme of Personalization' as examples to depict certain classes of personalization constraints. Most of the cases showed more than one limitation in applying personalization activities but the use of each case in the classification scheme was restricted to one to overcome redundancy and provide as much insight into the case study analysis as possible.

TABLE VI. EXAMPLES OF ANALYZED CASES

No.	Case	Constraint	Description
1	Amazon	Internal technological of matchmaking	Recommends products based on transactions and other data
3	Netflix	External organizational of delivery	Recommends films based on transactions and other data
5	Pandora	External technological of matchmaking	Recommends music songs based on users' music taste
9	Travelocity	External technological of delivery	Provides personalized services (flight status) for booked flights
12	Foursquare	External organizational of matchmaking	Shows and recommends 'places' nearby based on friends' taste
14	Southwest	Internal	Submits special offers based

No.	Case	Constraint	Description
	Airlines	organizational of matchmaking	on a defined home airport
16	Google Search	Internal technological of data collection	Provides personalized search results based on prev. searches
17	Google AdSense	External technological of data collection	Provides personalized ads based on users' preferences and sites
18	Facebook Sponsored Stories	Internal organizational of data collection	Adds information of companies to status reports
20	L'Oréal	Internal organizational of delivery	Personalizes web sites to provide individual country information
22	Erste Group Bank	External organizational of data collection	Provides personalized banking channels based on preferences
23	Direktanlage	Internal technological of delivery	Provides different tools based on customers' choice

By analyzing the cases, the 'Classification Scheme Personalization Constraints' can be depicted as shown in Table VII. It is divided into three dimensions which are origin (internal, external), subject (technological, organizational) and time (data collection, matchmaking, delivery).

The origin of personalization constraints can be either external or internal. If an internal personalization constraint is discovered, companies are normally able to resolve these issues internally. The management of such constraints is often time consuming and requires specific technological know-how, but the company has the ability to control the solution. On the other side, external constraints of personalization are based on factors which the company is not able to influence significantly. These barriers include strict regulations in data security or licensing, privacy concerns of customers or the active participations of users.

Subsequently, the subject of the personalization constraint can be divided into technological and organizational. Technological subjects include the development of accurate mathematical recommendation algorithms or the implementation of data collection systems. Organizational constraints on the other side include management issues like the registration of new users, the control of misuse as well as the resolution of timing matters.

Finally, the last dimension of personalization constraints covers the time on which the obstacle occurs. The time-dimension is mainly derived from the personalization process [11] [16] and includes data collection, matchmaking and delivery. Constraints of data collection include the inability of a system to gather user data which could be used for personalization issues. Matchmaking constraints refer to problems of the system or the organization to find products, services or results in general which fit the individual needs and preferences of the user. Delivery issues are then personalization constraints which occur during the delivery phase of the results. For example if the user is not able to retrieve his personalized results because of licensing regulations.

TABLE VII. CLASSIFICATION SCHEME OF PERSONALIZATION CONSTRAINTS

		data collection	match-making	delivery
technological	**internal**	Case 16	Case 1	Case 23
	external	Case 17	Case 5	Case 9
organizational	**internal**	Case 18	Case 14	Case 20
	external	Case 22	Case 12	Case 3

After defining the different constraints of personalization activities finally the following 'Standard Types of Personalization Environments' in marketing were derived from the case study analysis (see Table VIII). They are divided by the origin of possible personalization constraints. Within each group, a differentiation between a high and low impact of the constraint on the personalization activity can be made. The four standard types are named 'Risk', 'Performance', 'Dependence' and 'Flow'.

TABLE VIII. STANDARD TYPES OF PERSONALIZATION ENVIRONMENTS

		external constraints	
		high	**low**
internal constraints	**high**	Risk	Performance
	low	Dependence	Flow

If a business is conducting personalization activities in the 'Flow' environment, external and internal constraints have a low impact on these activities. One example of this standard type is case no. 14 'Southwest Airlines'. They offer a basic form of personalization by providing special offers to registrated customers based on their chosen home airport. Internal technological constraints are low, because the company only needs to match new offers to the explicit choice and send a message to the customer. Internal organizational constraints lie in the recurring provision of special offers for all available airports. External technological constraints include the availability of the customers' device. External organizational constraints are mainly covering privacy and timing issues. Summarizing the results, businesses which are operating in this environment can easily perform their personalization activities, because they do not face severe obstacles. They should soon end up in a 'flow' where personalization activities are conducted on a regular basis without any serious negative feedback.

In the 'Performance' standard type, businesses are facing high internal constraints, while external constraints on planned personalization activities are low. One conducted case study which could match with this type is case no. 16. Google Search personalizes their search results for each customer by using a vast basket of data. Constraints to this type of personalization are mostly internal technological barriers. Once Google Search developed a suitable algorithm to match individual preferences and search results, they were able to personalize their offerings. External constraints are typically low in this type and also Google Search is not confronted with many external barriers except some privacy issues and negative feedback on the constriction of search results. The standard type is named ‘Performance’ because the individual performance of the company is needed to successfully conduct personalization activities.

‘Dependence’ is the third standard type and depicts business situations where companies are facing high external constraints of personalization activities but low internal constraints. Foursquare (case no. 12) is one example of this standard type. Foursquare is highly dependent on external help for the provision of the personalization. Users need to overcome their privacy concerns, provide their individual position and add new places to the system. Once these external barriers are overcome is it easily manageable for the company to provide them their very personal results based on their current position. There are thus only limited internal constraints. Due to these reasons, the standard type is named ‘Dependence’.

The last type is named ‘Risk’ because personalization activities in this section are facing high internal as well as high external constraints. One example for a situation where businesses are working in the risk standard type is the banking sector. In case no. 22 (Erste Group Bank) the bank needs to overcome internal constraints like the development of applications for different operating systems of mobile devices, even though this is not their core competence. And they also need to manage external constraints which mainly include privacy concerns of the customers when they are enabled to use new communication channels as well as security issues. This standard type is named risk, because businesses need to work hard to ensure the management of the internal and external obstacles which are related to these personalization activities.

V. Conclusion, Limitations And Future Research

Beside the development of an overview of popular personalization issues of the last ten years and a ‘Morphological Box of Personalization’, the ‘Classification Scheme of Personalization Constraints’ as well as four ‘Standard Types of Personalization Environments’ have been introduced. The classification scheme is based on three dimensions, namely origin (internal, external), subject (technological, organizational) and time (data collection, matchmaking, delivery). The standard types are named ‘Risk’, ‘Performance’, ‘Dependence’ and ‘Flow’ and represent four personalization environments where external and internal personalization constraints are either high or low. The results provide a general starting point for businesses to concern themself with possible obstacles when planning personalization activities. They also enhance the existing personalization literature by introducing a general theory of limitations and constraints of personalization activities in digital environments.

Limitations of the study include on the one hand the choice of the scientific database on which the scientometric approach for theory development is based. Naturally, databases are not able to cover all journal titles. On the other hand the used multi case study research approach to verify the proposed model is only able to give an analytical but not statistical generalization of the research issue.

Further research will be conducted to describe the observed limitations and constraints of personalization activities in more detail. Furthermore, a proposition of success factors which lead to the adequate realization of personalization activities needs to be depicted. Finally, the consequences as well as risks and opportunities of personalization activities and their constraints on corporate communication processes as well as marketing will be analyzed and highlighted in future research.

References

[1] P. Schönhagen, „Gesellschaftliche Kommunikation im Wandel der Geschichte,“ in Medienpsychologie, B. Batinic, and M. Appel, Eds. Heidelberg: Springer, 2008, pp. 46-76.

[2] M. Hilbert, and P. López, “The World’s Technological Capacity to Store, Communicate, and Compute Information,” Science, vol. 332, no. 6025, pp. 60-65, 2011.

[3] T. Davenport, L. Mule and J. Lucker, “Know What Your Customers Want Before They Do,” Harvard Business Review, December 2011.

[4] A. Ansarai, and C. Mela, “E-customization,” Journal of Marketing Research, vol. XL, pp. 131-145, 2003.

[5] M. Porter, Competitive advantage: creating and sustaining superior performance. New York: Free Press, 1985.

[6] K. Tam, and S. Ho, “Understanding the Impact of Web Personalization on User Information Processing and Decision Outcomes,” MIS Quarterly, vol. 30, no. 4, pp. 865-890, 2006.

[7] D. Lee, J. Ahn, and Y. Bang, “Managing Consumer Privacy Concerns in Personalization,” MIS Quarterly, vol. 35, no. 2, pp. 423-444, 2011.

[8] H. Fan, and M.S. Poole, “What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems,” Journal of Organizational Computing and Electronic Commerce, vol. 16, no. 3&4, pp. 179-202, 2006.

[9] A. Sunikka, and J. Bragge, “Applying text-mining to personalization and customization research literature – Who, what and where?,” Expert Systems with Applications, vol. 39, no. 11, pp. 10049-10058, 2012.

[10] A. Sunikka, and J. Bragge, “What, Who and Where: Insights into Personalization,” Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008), pp. 283-292, 2008.

[11] J. Vesanen, and M. Raulas, “Building Bridges for Personalization: A Process Model for Marketing,” Journal of Interactive Marketing, vol. 20, no. 1, pp. 5-20, 2006.

[12] N. Tintarev, and J. Masthoff, “Evaluating the effectiveness of explanations for recommender systems – Methodological issues and empirical studies on the impact of personalization,” User Modeling and User-Adapted Interaction, vol. 22, pp. 399-439, 2012.

[13] S. Ho, “The attraction of internet personalization to web users,” Electronic Markets, vol. 16, no. 1, pp. 41-50, 2006.

[14] N. Arora, X. Dreze, A. Ghose, J.D. Hess, R. Iyengar, B. Jing, et al., “Putting one-to-one marketing to work: Personalization, customization, and choice,” Marketing Letters, vol. 19, no 3-4, pp. 305-321, 2008.

[15] A. Montgomery, and M. Smith, “Prospects of Personalization on the Internet,” Journal of Interactive Marketing, vol. 23, no. 2, pp. 130-137, 2009.

[16] G. Adomavicius, and A. Tuzhilin, “Personalization Technologies: A Process-Oriented Perspective,” Communications of the ACM, vol. 48, no. 10, pp. 83-90, 2005.

[17] K. Kwon, and C. Kim, “How to design personalization in context of customer retention: Who personalizes what and to what extend?,” Electronic Commerce Research and Applications, no. 11, pp. 101-116, 2012.

[18] G. Miceli, F. Ricotta, and M. Costabile, “Customizing customization: A conceptual framework for interactive personalization,” Journal of Interactive Marketing, vol. 21, no. 2, pp. 6-25, 2007.

[19] K. Kalyanam, and S. McIntyre, "The e-marketing mix: A contribution of the e-tailing wars," Journal of the Academy of Marketing Science, vol. 30, no. 4, pp. 487-499, 2002.

[20] F. Zwicky, Entdecken, Erfinden, Forschen im Morphologischen Weltbild. Munich: Droemer Knaur, 1966.

[21] R. Yin, Case study research. Beverly Hills: Sage Publications, 1984.

[22] K. Eisenhardt, "Building Theories from Case Study Research," The Academy of Management Review, vol. 14, no. 4, pp. 532-550, 1989.

[23] K. Pousttchi, and D. Wiedemann, "A Contribution to Theory Building for Mobile Marketing: Categorizing Mobile Marketing Campaigns through Case Study Research," International Conference on Mobile Business (ICMB), pp. 1-8, 2006.

[24] B. Murthi, and S. Sarkar, "The Role of the Management Sciences in Research on Personalization," Management Science, vol. 49, no. 10, pp. 1344-1362, 2003.

[25] P. Schubert, and M. Koch, "The Power of Personalization: Customer Collaboration and Virtual Communities," Proceedings of the 8th Americas Conference on Information Systems, pp. 1953-1965, 2002.

[26] K. Kwon, J. Cho, and Y. Park, "How to best characterize the personalization construct for e-services," Expert Systems with Applications, vol. 37, pp. 2232-2240, 2010.

[27] P. Schubert, Virtuelle Transaktionsgemeinschaften im Electronic Commerce: Management, Marketing und soziale Umwelt. Cologne: Josef Eul, 1999.

[28] K. Riemer, and F. Brüggemann, „Personalisierung der Internetsuche – Lösungstechniken und Marktüberblick," Wirtschaftsinformatik, vol. 49, no. 2 pp. 116-126, 2007.

[29] M. Kunczik, and A. Zipfel, Publizistik: Ein Studienhandbuch. Stuttgart: UTB, 2005.

[30] W. Schweiger, Theorien der Mediennutzung. Wiesbaden: VS, 2007.

[31] R. Baraglia, and F. Silvestri, "Dynamic Personalization of Web Sites Without User Intervention," Communications of the ACM, vol. 50, no. 2, pp. 63-67, 2007.

Formal Method to Derive Interoperability Requirements and Guarantees

Hazem El-Gendy,
Ph.D., P.Eng.
Faculty of CS &IT Ahram Canadian University

Magdi Amer,
Ph. D., P. Eng.
Faculty of Eng., Oum Kora Univ.,
Makka, Saudi Arabia,

Ihab Talkhan,
Ph. D., P. Eng.
Faculty of Eng, Cairo Univ.,
Giza, Egypt.

***Abstract*— Interoperability among telecommunications systems, possibly by different vendors, is essential for both the development of many telecommunications networks, and today's civilization development. Interoperability testing is very costly, as it has a complexity of (n**2) for n systems, and somewhat informal. In this paper, we develop a 'Conformance Testing (CT)'-based formal technique to determine interoperability requirements/guarantees. It allows automated derivation of the interoperability' requirements of various networks as well as the interoperability guarantees among different telecommunications systems. This is achieved using static analysis of the conformance classes of the standard and knowledge of the implementation's degree of conformance (DoC) of the telecommunications systems. Consequently, it results in a lot of cost saving in addition to being a formal technique.**

Keywords—Computer/Communications Protocols and Standards; Conformance Requirements and Classe; Interoperability; Protocol Data Units (PDUs); Capabilities.

I. Introduction

Different telecommunications networks are typically by different vendors. This is often associated with many interoperability problems. This is the case even when these products are conforming products to an international standard. Also, different telecommunications products by the same vendor and that play different roles in the network may have interoperability problems in certain cases. Nevertheless, building a multi-vendor network has become a key requirement of every intelligent telecommunications user. This is to relief that user from dependency on any single telecommunications vendor. Such independence, when achieved, results in better economics of the network expansion and more importantly a feature richer network and continuous supply of telecommunications products.

This motivated the work on Interoperability Testing. Methods [8.9.10,11]. Interoperability testing methods developed so far are mainly informal methods that are both protocol dependent and product dependent. By protocol dependent, we mean that the method's applicability is limited to only one protocol. Product dependency means that the method has to be applied for every pair of products for which interoperability is required; consequently, the complexity of the "interoperability testing"-based methods for interoperability analysis is (n**2). So, for n different telecommunications products, all the following are required:

- Designing of n * n Interoperability Test Suites;
- Running the n * n Interoperability Tests Suites; and
- Conducting n * n Interoperability Test Results Analysis.

In this paper, we develop a CT-based method for interoperability analysis and guarantees. CT is the type of testing that aims at increasing the confidence in the correct implementation of the telecommunications products. As CT is also an ISO (which aims at facilitating Open Systems Interconnections) research work, it has to also increase the confidence that conforming implementations interoperate. The CT-based method has a complexity of just n for n different products. As the method depends on static analysis of the standard and the determination of the DoC of every product, no additional tests are required: this includes no design at all of any interoperability test suites. Furthermore, the method is formal and facilitates full automation.

II. Testability-Directed Specification Of A Protocol

A Testability-Directed Standard of a communications Protocol (TDSP) [1] is a tuple:

TDSP := (P, I, BS, TDPICSP & ConfStat, ServDesc, NCTC) where:

P: set of service parameters.

I: set of interactions

BS: Specification of allowed protocol behavior. From BS, all the allowed sequences of interactions (BSeq) can be derived. Also, the inter-dependencies between the interaction parameters can be extracted. Each sequence (trace) $BSeq_s = (s, P_s, C_s)$ represents:

a) A syntactically allowed ordering s of interactions.

b) A number of constraints (C_s) on the parameters (P_s) of the interactions in the ordering s. From BS, it is possible (in principle) to derive all possible traces (BSeq) as follows:

Let

- S be the set of syntactically allowed orderings of interactions as derived from BS;

- R : Constraints relation

- For each $s \in S$, $\exists\ C_s \subseteq R$ where C_s represents a set of constraints on parameters P_s in s.

Then, the set of all possible traces is denoted BSeq = $\{(s.\ P_s.\ C_s)\}$ where

$$s\ (=<i,.i_2.. .\ ,i_n>) \varepsilon\ S$$

$$P_s = \{(p_j,\ v\ (p_j)/\ \exists\ i(p_1,p_{2,}..\ .\ ,p_k) \in s,\ K \geq j \geq 1\}$$

$$C_s=\{(p_i.p_j)/p_i,p_j\ \varepsilon\ P_s\}$$

TDP1CSP: Testability-Directed Protocol's Implementation Confomance Statement Proforma [l].

ConfStat: Conformance Statement

ServDesc: Service Description of the various capabilities covered by the standard.

NCTC: Non-Conformance Testing Clause.

Meas of Specifying Components:

I, P, and BS can be formally specified using an FDT. TDPICSP is given in tabular form. ServDesc, ConfStat, and NCTC are given in a natural language.

III. Conformance Classes

A Standard typically has capabilities whose faithful support is mandatory for conformance or optional or conditional [2.6.7]. This generates Conformance Classes. Thus, we have: Let M, 0, and C be the sets of mandatory, optional, and conditional capabilities respectively. Let also, R: C x C. *Interdependency Relation* where $(c_i,\ c_j)$ is in R iff faithful support of c_i requires faithful support of c_j. Then, ConfClass is defined to be Set of Conformance Classes =(MXOj) U DOj where Oj $\subseteq$ O and DOj = $\{c_i\ /(c_l,c_i) \in R$ and $c_l \in Oj\}$.

Every conformance class corresponds to a *self-contained* set of capabilities. Each conformance class identifies a unique set of conformance requirements that has to be satisfied by an implementation to be a conforming implementation to the conformance class.

Conforming implementations are more likely to perform the required functionality as well as interoperate than non-conforming implementations.

Interoperability problems typically involve:

- receiving an illegal PDU that the implementation cannot understand; or
- receiving a legal PDU but with an illegal parameter; or
- receiving a legal PDU in an unexpected state.

These problems may result from error in designing the protocol itself or in implementing the protocol. Protocol Verification and Validation aim is to resolve errors in designing the protocols. Conformance Testing may handle errors (uncover them) in implementing the protocols.

IV. Interoperability Requirements And Guarantees

In this section, we develop a formal method to derive interoperability requirements and guarantees. The method is based on conformance testing and the use of a Testability-Directed protocol Standard.

A conformance class ConfClassj is *interworkable* with a conformance class ConfClass;. denoted by ConfClassj INTERW ConfClassj. iff ConfClass_i meets all the requirements for interworking with ConfClass_j; i.e.. ConfClassj may not receive, from ConfClass_j a request for a service for which ConfClassj cannot guarantee delivery as per the standard. However, ConfClassj may offer more services functions/capabilities than ConfClass_j. It is important to note that the relation INTERW is not symmetrical as anticipated. "Conformance Testing" of an IUT determines the IUT's "Degree of Conformance (DoC)".

Here, we consider the lUT's (maximum) degree of conformance to correspond to the largest conformance class that the IUT faithfully supports. The determination of the Conformance Classes and the DoCs facilitates static investigation and study of the interoperability between the implementations of the various conformance classes: this is illustrated in the next section. This is particularly important because it provides extensive information about the potential for interoperability between the various conforming systems without having to have any physical development of any of these systems. Such information can assist manufacturers in making decisions (marketing decisions) about what standard capabilities to support, and assist the users in making decisions on what capabilities to require faithful support for in the products they intend to purchase.

Interoperability among implementations: An implementation I is said to be[1] interoperable with an implementation I', denoted by I INTERO I', iff there is not a capability, out of those offered by the standard, that I' may request from I and I cannot faithfully offer.

Lemma: For two implementations I and I': I INTERO I' iff I CONF ConfClass_I, I' CONF ConfClass_j, and ConfClass_I INTERW $\text{ConfClass}_{I'}$.

Proof: Follows from having I CONF ConfClass_j, I' CONF $\text{ConfClass}_{I'}$ and the definition of INTERW.

V. Example

Automated derivation of interoperability requirements and guarantees are illustrated by a Testability-Directed version of the Transport Layer Class 2 protocol given as follows.

BEGIN_TDPICS_PROFORMA_TP(2)
Identifiers:

Supplier	
Siandard Version	15 8073 V 1
Date of Statement	
Implementation Identification	
Extra Information for Testing	

CAPABILITIES:

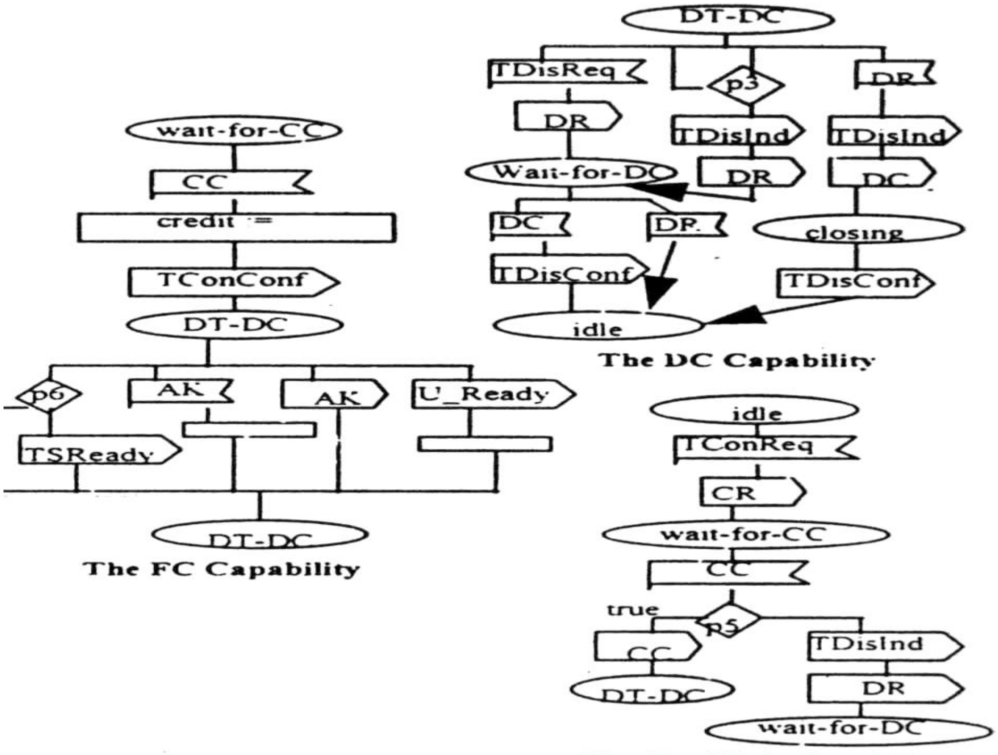
#	Capability	Ref	Status	Su
A1	Initiating the establishment of a connection	IC	O_{IR}.1	
A2	Responding to a request to a connection	RC	O_{IR}.1	
A3	Transferring Data	DT	M_{iR}	
A4	Disconnection of a connection	DC	M_{iR}	
A5	Flow Control	FC	A1:O_{iR}	
A6	Initiator negotiation of options	N_I	A1:O_I	
A7	Responder negotiation of options	Nr	A2:M_R	

PDUS/SERVICEPRIMITIVES:

£	PDt	Ref	Status	Su
Bl	Output CR	CR	(al v a6): M_I	
B2	Input CR	CR	(a2 v a7) M_R	
B3	Output CC	CC	a2M_I	
B4	Input CC	CC	(al v a5 v a6):M_R	
B5	Input DR	DR	(al v a4) M_R	
B6	Output DR	DR	(*a2* v a4 v a6):M_I	
B7	TransportData	DT	a3.M_{IR}	
B8	Disconnection	DC	(a4 va7):M_R	
B9	Acknowledgement	AK	A5:M_{IR}	
B10	TConnectionIndic	TCI	(A2 v A7):M_I	
B11	TConnectionRespon	tcr	a2:M_R	
B12	TDisconnectionReq	tdr	(a2 v a4):M_R	
B13	TDisconnectionConf	tdc	A4:M_I	
B14	TConnectionConfin	tcc	(al v a5):M_I	
B15	TConnectionRequest	tcr	(al v a6):M_R	
B16	TDisconnectionIndic	tdi	(al v a4 va6): M_I	
B17	User_Ready	UR	A5:M_R	
B18	TSReady	tsr	A5:M_I	

PARAMETERS:

#	Parameter	Ref	Status	Su
C1	Proposed_options	B10, B15	(A6 v A7):M	
C2	Accepted_options	Bl1, Bl4	(A6 v A7):M	
C3	Options_ind	Bl, B2, B3, B4	(A6 v A7) M	
C4	Credits	Bl,B2,B3,B4, B9	A5:M	
C5	Credit_value	B17	A5:M	
C6	User_data	B2, B3, B4, B5, B7, B8, B14	M_{1R}	



The DC Capability

The FC Capability

The NegOInit Capability

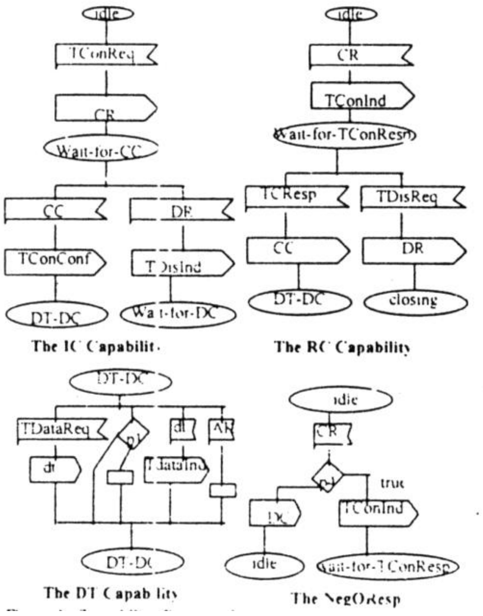


Figure 1: TD-BS of TP(2)

The Conformance Statement component of this testabiliry-directed standard is now constructed according to the guidelines in [1.6,7]. and the International Conformance Testing Standard ISO/IEC 9646.

END_Tcstability-Directed_TP_Class_2_Standard

Analysis with Respect to Interoperability Requirements/ Guarantees

The TDPICSP-based conformance requirements, as indicated by the TDPICSP. indicate that every implementation, to conform to the specification, has to support at least three capabilities: DT. DC. and either IC or RC; also, if the RC capability is supported, the support of the NegO$_R$ is mandatory. A conforming implementation supports up to seven capabilities. The support of the NegO$_I$ (NecO$_r$) capability requires the support of the IC (RC) capability. Consequently, there are nine conformance classes of implementations as follows (where the # field provides the conformance class identifier and the other field provides the capabilities that constitutes the conformance class):

CONFORMANCE CLASSES OF TP(2)

#	Capabilities of the cc#	#	Capabilities of the cc#
1	IC,DT, and DC	2	IC, DT, DC, and NegO$_I$
3	IC,DT, DC, and FC	4	RC, DT, DC, andNegO$_R$
5	IC, DT, DC, Neg Oj, FC	6	IC,RC,DT, DC, FC,NegO$_R$
7	IC, RC, DT, DC, Neg O$_r$	8	IC,RC,DT,DC,NegO$_R$, NegO$_I$
9	IC, RC, DT, DC, FC, Neg O$_r$, Neg Oj		

Every conformance class corresponds to a unique degree of conformance of implementations; consequently, there are ten implementation's degrees of conformance (IDoC) (one degree of conformance corresponding to each conforming implementation class plus a zero value (Implementation's Degree of Conformance of Zero) corresponding to implementations that fail to faithfully support any of the mandatory requirements). The static analysis of these conformance classes along with the "Implementation's Degree of Conformance" of the various implementations provides a good basis for determining the chance of interoperability between the various conforming implementations: for example. two conforming implementations with Implementations Degree of Conformance = 4 (i.e.. faithfully supports conformance class 4) cannot interwork because neither of the two implementations can initiate the establishment of a connection (each of the two acts always as a responder); while two implementations with Implementation Degree of Conformance = 9 can interwork with each other. Such type of interworking, that is capability-based, is called, here, "Capability (functional) Interworking" and can be determined by static analysis of the testability-directed standard.

On the other hand, class-9-conforning implementations (implementations that faithfuly) supports every capability, in the standard) have the highest chance to successfully interwork as long as TP(2) standard is considered, with every other conforming implementation. Generally, these class 9 conforming implementations interwork with every other faithfully conforming implementation provided that the standard is error free: such interworking covers only those layers/protocols covered by the standard.

For the testability-directed standard. We have the following interworking guarantees, illustrated in Figure 3 in a lattice form, where every node represents implementations that faithfully support a valid class and the conforming implementations of a node can interwork with every conforming implementation in its sub-tree: the type o! interworking meant here is that the conforming implementations of a node may not initiate/invoke a capability that the conforming implementations of any of its parent nodes cannot positively (according to the standard) respond to it. For example, the IC capability of a conforming implementation interworks with an RC capability of another conforming implementation: consequently, implementations conforming to class 1 interwork with the implementations conforming to class 4 but not class 1: consequently, a class -4 node (but not a class 1 node) is a parent of a class 1 node. Also, class-4-conforming implementations can intenvork with class-8-conforming implementations because the latter may not initiate the invocation of a capability that the class-4-conforming implementations cannot faithfully respond to it (i.e., there is a class 4 node that is a parent to a class 8 node).

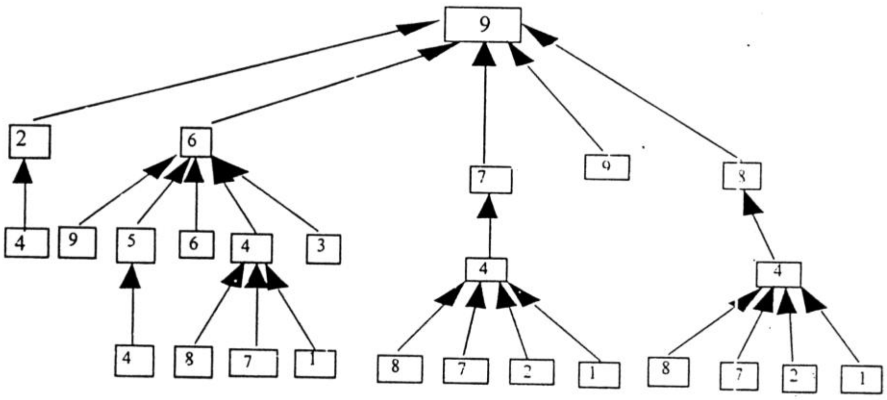


Figure 2: Lattice Representation of the interoperability guarantees between the various conforming classes.

The "interwork with" relation as defined on the conformance classes is not reflexive because, for example, the class-1-conforming implementations cannot interwork with the class-1-conforming implementations (they deadlock). Also, "intenvork with" relation is not transitive. For example, the class-1-conforming implementations interwork with class-4-conforming implementations and class-4-conforming Implementations interwork with the class-2-conforming implementations. But, class-1-conforming implementations do not interwork with class-2-conforming implementations.

However, a testability-directed standard facilitates the static analysis of the specification for interworking guarantees based on the implementation's degree of conformance which, in turn, is very useful for resolving many conformance testing issues and interoperability issues as well as assisting industry in making intelligent decisions regarding what capabilities to require the implementations to faithfully support.

VI. Conclusions

In this paper, we have developed a CT-based technique to analyize interoperabiliy requirements and guarantees. The technique is a formal technique which makes it not error prone. Also, the technique *is* general enough to be applied to various communications protocols as far as they are formally specified using an International FDT: Lotos, Estelle, or SDL. The analysis is a static analysis which saves a lot of cost in performing interoperability testing. For n implementations of a standard, the method requires running only n conformance tests rather than n times n interoperability tests. Furthermore, the n conformance tests are needed in any way. Consequently, the method does not require conducting any additional tests and therefore saves a lot of cost. The applicability and practicality of the method has been demonstrated by a real large protocol standard. Finally, the method can be fully automated.

References

[1] Hazem El-Gendy, Magdi Amer, and Osman Ibrahim, “Towards Modeling of Communications Protocols“, IEEE 20th Telecommunications Forum (TELFOR), Nov. 20-22, 2012, Belgrade, Serbia, pp.95-99.

[2] Hazem El-Gindy. "Towards a Foundation for Conformance Testing Requirements: FDT-Based CTRs". Proceedings of th; IEEE International Conference on Electronics. Circuit;, and Systems. Rodous. Greece, 16-18 Sept. 1996. Wa[1] also accepted for Publication in the Proceedings of the International Conference on Networks sponsored by the nternational Association of Science and Technology for Development (IASTED). Or'ando. Florida. USA. January 8-10. 1996.

[3] Hazem El-Gendy. "A New Theory for Equivalence of Process Specifications". Proceedings of the IEEE International Conference on Electronics. Circuits, and Systems. Rodous. Greece. 16-18 Sept. 1996. Was also accepted for Publication in the Proceedings of the International Conference on Networks sponsored by the International Association of Science and Technology for Development (IASTED). Orlando. Florida. USA. January 8-10,1996.

[4] Hazem El-Gendy, "A New Method for Deriving Test Sequences For Protocols Specified in LOTOS". Proceedings of the International Conference on Networks sponsored by the International Association of Science and Technology' for Development (IASTED), Orlando. Florida. USA. January 8-10. 1996.

[5] Hazem El-Gendy. and Osman Abou-Rabia. "Automated Derivation of Test Sequences for Protocols and Software Specified in Lotos", Proceedings of the IEEE 2nd International Conference on Electronics, Circuits, and Systems. Amman. Jordan. December 18-21. 1995. pp. 220-225.

[6] Hazem El-Gendy and Robert Probert. "Relevant Testability Aspects of a Communications Protocol Standard". Proceedings of the 4th IEEE Maylaysia International Conference on Communications (IEEE M1CC95). Langkawi Island. Nov. 20-23, 1995. pp. 14.2.1-14.2.6.

[7] Hazem El-Gendy and Robert Probert. "Towards a Framework and Methodology for Conformance Testing Requirements". Proceedings of the 4th IEEE Maylaysia International Conference on Communications (IEEE MICC'95). Langkawi Island. Nov. 20-23. 1995, pp. 14.1.1 -14.1.8.

[8] Stephen Castro. "The Relationship Between Conformance Testing of and Interoperability- Between OS1 Systems". Computer Standards & Interfaces, Vol. 12. No. 1. 1991. pp. 3-11.

[9] Jay Gadre. Chris Rohrer, Catherine Summers, and Susan Symington, "A COS Study of OSI Interoperability". Computer Standards & Interfaces. Vol. 9. No. 3, 1989. pp. 217-237.

[10] L. Legrand and G. Cesarine, "Network Management Interoperability Testing". Preceding of the llth IFlP's International Workshop on Protocol Specification. Testing, and Verification. 1991.

[11] Pierre de Saqui-Sannes. Jean-Pierre Courtiat. and Rene Casadessus. "Verification by Abstraction as a Preamble for Interoperability Test Suite Generation". Proceedings of the 14th IFIP WG6.1 International Symposium on Protocol Specification. Testing and Verification. Vancouver. BC. Canada. June 1994. pp. 152-166.

[12] Mohamed Wagdi Youssef and Hazem El-Gendy, “Scrampling and Encrypting-Based Authontication for for Open Networks Communications”, International Journal of Computer Science and Network Security, June 2011, pp. 24-29.

[13] Hazem El-Gendy, Nabil El-Zant El-Kadhi, Osman Ibrahim, and Narayan Debnath, “Complete Management Scheme for Intelligent Terminals for Information Super-highways and Its Design”, International Journal of Computational Methods in Sciences and Engineering, Volume 10, No.s 1,2, 2010, pp. 247-257.

[14] Nabil El-Kadhi and Hazem El-Gendy, "An Intelligent Bidirectional Authentication Method", The International Journal on Computer and Network Security, published in Vienna, Austria, 2010, pp.64-70.

[15] Hazem El-Gendy, Nahla El Zant El Kadhi, Osman Ibrahim, and Narayan Debnath, "Complete Management Scheme for Intelligent Terminals for the Information Super-highways and its design", Journal of Computational Methods in Sciences and Engineering, Vol. 10, Number 1,2, 2010, IOS Press, Netherland.

[16] Hazem El_Gendy, "Towards ISO Interface Protocol for Interactive Multimedia Intelligent Terminals", Proceedings of the 9th International Conference on Computer Systems and Applications, Hammamet Tunisia, May 16-19, 2010 sponsored by ACS, IEEE, and IEEE Computer Society.

[17] Hazem El-Gendy, Osman Ibrahim, Nabil El Kadhi, and Narayan Debnath, "Management Scheme for Intelligent Terminals for the Information Super-highways", Proceedings of the 9th International Conference on Computer Systems and Applications, Hammamet Tunisia, May 16-19, 2010 sponsored by ACS, IEEE, and IEEE Computer Society.

[18] Hazem El-Gendy, "Formal Development of Universal Protocol Implementations Conformance Statement Proforma", Proceedings of the 7th ACS/IEEE International Conference on Computer Systems and Applications sponsored by ACS, IEEE, and IEEE Computer Society, Rabat, Morocco, May 10-13, 2009.

[19] Hazem El-Gendy, Mohamed Gebriel, Ahmed Samir, Narrayan Debnath, and Nabil El-Kadhi, "Towards Mosques Management Information System", Accepted for Publications in the 7th ACS/IEEE International Conference on Computer Systems and Applications sponsored by ACS, IEEE, and IEEE Computer Society, Rabat, Morocco, May 10-13, 2009.

[20] Hazem El-Gendy, Nabil ElKadhi, and Narrayan Debnath, "Towards Sound Development of PIXITP, Conformance Test Suites, and Conforming Implementations For Various Formal Description Techniques", Proceedings of the 12th IEEE International Symposium on Computers & Communications sponsored by both IEEE Computer Society and IEEE Communications Society, Morocco, July 6-9, 2008.

[21] Hazem El-Gendy, Nabil ElKadhi, and Narrayan Debnath, "Formal Automated Transformation of SDL Specifications to Lotos Specifications", Proceedings of the 12th IEEE International Symposium on Computers & Communications sponsored by both IEEE Computer Society and IEEE Communications Society, Morocco, July 6-9, 2008.

[22] Hazem El-Gendy, Nabil ElKadhi, and Narayan Debnath, "Formal Automated Transformation of SDL Specifications to Estelle Specifications", Proceedings of the 23rd International Conference on Computers And Their Applications (CATA'08) sponsored by the International Society for Computers & their Applications (ISCA), April 8-10, 2008, Cancun, Mexico.

[23] Hazem El-Gendy, "Study of the Characteristics of CT-Equivalence with Proves", Journal of Computational Methods in Sciences and Engineering, Volume 6, Numbers 5, 6, 2006, pp. 171-179.

[24] Nabil El-Kadhi and Hazem El-Gendy, "Advanced Method for Cryptographic Protocol Verification", Journal of Computational Methods in Sciences and Engineering, Volume 6, Numbers 5, 6, 2006, pp. 109-119.

Stable Haptic Rendering for Physics Engines Using Inter-Process Communication and Remote Virtual Coupling

Xue-Jian He, Kup-Sze Choi
Centre for Integrative Digital Health, School of Nursing
The Hong Kong Polytechnic University
Hong Kong, PR

***Abstract*— Availability of physics engines has significantly reduced the effort required to develop interactive applications concerning the simulation of physical world. However, it becomes a problem when kinesthetic feedback is needed in the applications since the incorporation of haptic rendering is non-trivial, where fast haptic data update is demanded for stable rendering. In the regard, a framework for integrating haptic rendering into physics simulation engines is proposed. It mediates the update-rate disparity between haptic rendering and physics simulation engine by means of inter-process communication and remote virtual coupling, which fully decouples haptic rendering from complex physical simulation. Experimental results demonstrate that this framework can guarantee fast haptic rendering at 1k Hz even the physical simulation system operates at very low update rate. The remote virtual coupling algorithm shows better performance than the interpolation based methods in terms of stability and robustness.**

Keywords—haptic rendering; physics engine; inter-process communication; virtual coupling.

I. Introduction

To convey more immersive and realistic experiences in the cyberspace, many virtual reality (VR) systems have integrated kinesthetic feedback with the use of haptic interfaces. It has also been demonstrated that haptic sensation can greatly increase the effectiveness of the VR based simulation systems [1, 2]. With the availability of commercial haptic interfaces, software toolkits, and commercial haptics-enabled applications, haptics-enabled simulation technology is undergoing rapid and exciting growth [3]. However, realistic rendering of haptic sensation in interactive applications is a challenging task. On the one hand, real-time dynamics simulation and efficient collision detection are essential for modeling the physical world and graphics rendering, while high-performance haptic interfaces, stable force rendering algorithms, and a good understanding of the psychophysics of touch and perceptual factors [4] are required for modeling the forces resulting from the interactions in the simulated world.

Take virtual surgery applications as an example. It involves realistic replication of complex behaviors of the physical world, including interactions of surgical tools with multiple deformable objects, bleeding and smoke generation due to cautery procedures, as well as tissue approximation procedures such as suturing and stapling [5]. Although physics engines have been made available to facilitate virtual surgery by offering multi-physics simulation for rigid and deformable bodies, cloth and fluids, it required intensive computation and the high latency incurred is unfavorable for stable haptic rendering. Stable force feedback requires a data update rate of 1k Hz which is much more demanding than the refresh rate needed for smooth graphics rendering, i.e. 30~60 Hz [6]. The computational speed for realistic dynamics simulation of the physical world could even be much slower. It is therefore non-trivial and a trick task to integrate haptic rendering with physics engines, where a good compromise between the wide disparity in data update rate cannot be established easily. This presents a significant barrier for the incorporation of haptic rendering into real-time interactive virtual environments [5]. From a developer's point of view, the integration of haptics and dynamics simulation is a major consideration that would eventually affect the software quality and the development cycle [7].

At the programming level, one common solution to the issue of update rate disparity is to use multi-threading techniques, which allows haptic and graphic rendering to be executed concurrently in separate threads so that each rendering loop can run at its own refresh rate. While the responsiveness and throughput of the applications are improved, the multi-threading technique increases the complexity of software development where additional synchronization of shared resources is needed and the difficulty in debugging is increased. In particular, the error caused by one thread can kill the entire process because the entire memory space could be shared by all threads. Additional measures are thus needed for data synchronization during concurrent operations such as read-and-write memory access in the multi-threaded program. Otherwise, the data race can lead to deadlocks since all the threads are blocked and waiting for the others to release the required data.

An alternative approach to tackle the issue of update rate disparity is to parallelize the simulation processes so that they can operate as independent execution units that contain their own status information, use their own address spaces, while only interact with each other by means of inter-process communication (IPC) mechanisms. In the paper, this mechanism is adopted to develop a framework that uses the remote virtual coupling based on IPC in order to guarantee haptic rendering to execute at 1k Hz without being affected by physics simulation engines. This framework decouples haptic rendering from physics simulation, thus providing stable haptic feedback even under complex simulation scenarios running at a low update rate.

The rest of the paper is organized as follows. First, the related work of the study is reviewed in section II. The simulation framework is then presented in section III, including system architecture design, abstraction layers and haptic rendering algorithms. In section IV, the benchmark model is described to evaluate remote virtual coupling and interpolation based haptic rendering algorithms. The detailed implementation of the system and the evaluation are presented in section V. Lastly, discussions and conclusions are given in section VI and VII respectively.

II. Related Work

Haptic rendering refers to the process of computing and generating forces in response to user interactions with virtual objects [8]. The forces can be generated based on the penalty depth and coupling distance [9] which are calculated using the pose (position and orientation) of the haptic interface and that of the haptic pointer (rigid body, god-object [10], proxy [11]). To simulate the interactions with dynamic virtual worlds, it is straightforward to update the physics simulations at the same update rate of the haptic simulations. However, because of restrictions caused by computational resources and simulation complexity, it is difficult to ensure the haptic data to be updated at 1k Hz for stable force displaying [6].

To solve this issue, multi-rate techniques were proposed, which has succeeded in improving the stability and responsiveness of haptic rendering systems [12]. The idea was to perform updates of the virtual environment at a low frequency (limited by computational resources and system complexity) while performing high-frequency updates for force feedback by using simplified or approximate force models. These multi-rate techniques [13, 14, 24] were extended from the method of intermediate representation [15] which was proposed by Adachi et al. to enable interactions with a static virtual world by handling the simulation process with two separate threads, namely the collision detection thread and the haptic rendering thread. Each thread was executed at a different update rate and synchronized at the slowest update rate. In the slow collision detection thread, a plane was computed to serve as a unilateral constraint for the force-feedback thread. This technique was later adapted by Mark et al. to interpolate the intermediate representation between updates [14], which enabled higher stiffness values than the approaches that computed the feedback force values at the rate imposed by collision detection. Besides, Hasegawa et al. [13] made use of an impulse rendered in a haptic thread to update and stabilize movement of rigid bodies managed in a physics thread. The multi-rate approach remains a common approach for simulating haptic interactions with deformable models and has been further developed for many other applications [16~18].

Virtual coupling [9] is another robust stable haptic rendering algorithm, which was first proposed by Colgate et al. The algorithm overcomes the instability caused by the artificial addition of energy into the simulation system [19]. A damped spring was introduced between the pose of the haptic device and the simulated pose, where a trade-off between stability and performance was achieved by tuning the spring stiffness. Later, Akahane et al. [20] implemented a haptic display at 10k Hz update rate by interpolating and up-converting the forces generated by virtual coupling to achieve a stiff, high resolution haptic rendering system. Besides, Otaduy et al. [21] developed a haptics-enabled system with a haptic thread of high update rate and the contact thread of low update rate. The haptic thread then calculated the coupling force and simulated the dynamics of the haptic pointer to realize a stable haptic display with a low-mass-value haptic pointer.

To improve the virtual coupling method, a constraint-based coupling algorithm based on the god-object method [10] was proposed by Ortega et al. [22]. Unconstrained and constrained acceleration of haptic pointer were introduced to calculate the feedback forces. The constrained-based coupling technique can suppress the artificial friction or sticking effect occurred in the conventional virtual coupling. Glondu et al. [23] proposed an algorithm called haptic sub-world using a contact graph, which allowed the simulation of rigid bodies with high update rate and enabled the rendering of feedback forces without artifacts. However, this method cannot handle the situation where all rigid bodies are in contact with each other.

However, most of these multi-rate and virtual coupling haptic rendering techniques were implemented in the multi-threading framework, which has some intrinsic shortcomings as mentioned in the previous section. To solve these issues when integrating haptics with physics engines, a remote virtual coupling based on the IPC framework is proposed in this paper.

III. Simulation Framework

In this section, the framework proposed for stable haptic rendering in the integration of haptics and physics simulation is presented. The system architecture, abstraction layers, haptic rendering algorithms involved in the framework will be discussed.

A. Client-Server Architecture Design

By employing the client-server architecture design, the haptics and physics simulation are clearly specified by properly defining the communication protocol. The block diagram in Fig. 1 illustrates the client-server architecture designed for the framework. The server side hosts the physics simulation application that provides the simulation results by utilizing a physics engine. The client side hosts the haptic rendering application mainly responsible for calculating feedback forces at a 1k Hz update rate based on the simulation results transferred from the server. The communication layer operates on a protocol defined to ensure information such as poses of the haptic device and simulation results is exchanged timely and correctly. Notice that the proposed framework only emulates a network. It runs on a single computer without physical connection to other computers.

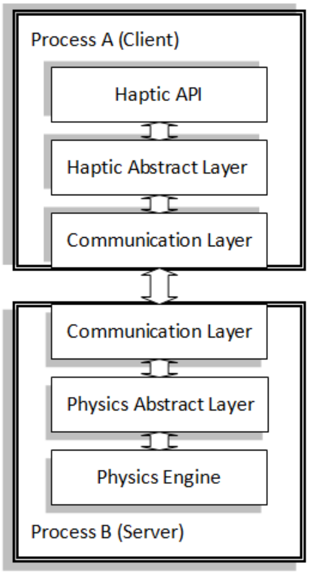


Figure 1. Block diagram of the system architecture design.

In the server side, instead of allowing the physics engine to interface directly with the communication layer, the physics abstraction layer, as an abstraction interface of the physics simulation Application Programming Interface (API), is introduced to improve development flexibility and ease of use.

Here, implementation details of the physics engine are abstracted to provide an organized and clean interface to the developers, where the abstraction layer enables developers to implement their version of physics simulation systems through a unique interface, thus allowing them to switch between different physics engines depending on their needs and performance level required.

In the client side, the Haptics Abstraction Layer (HAL) is an interface defining a common API that provides a robust toolkit for accessing functionalities of haptic devices from different manufacturers. It exposes a set of interfaces that hide the differences among the devices and makes the application device independent. This layer is highly extensible and customizable to cater for diverse types of haptic devices and special needs in individual applications.

The communication layer (CL) is responsible for exchanging information between the server and the client. As shown in Fig. 2, the information include haptic interface pose, velocity, button status, and so on from the client side, and the simulation results such as the virtual tool's proxy pose and velocity from the server side. The client CL runs on a thread at 50 Hz which parcels and sends the spatial information of the haptic device to the server CL.

The server CL will block its own thread until it receives one data packet. The data from the packet will then be extracted by server CL to calculate the force for the proxy in the simulation environment, based on the pose difference between the proxy and the haptic device using a spring-damper model. These events are also described in the sequence diagram in Fig. 3.

It is worth noting that the coordinates of the haptic space might be different from that of the simulation engine and coordinate mapping would then be required.

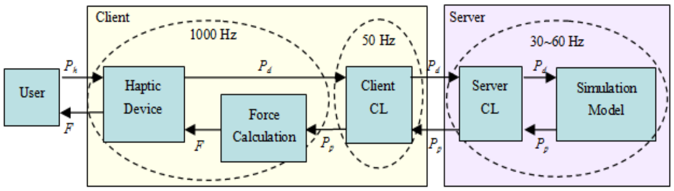


Figure 2.Data flow in the client-server based haptic rendering framework.

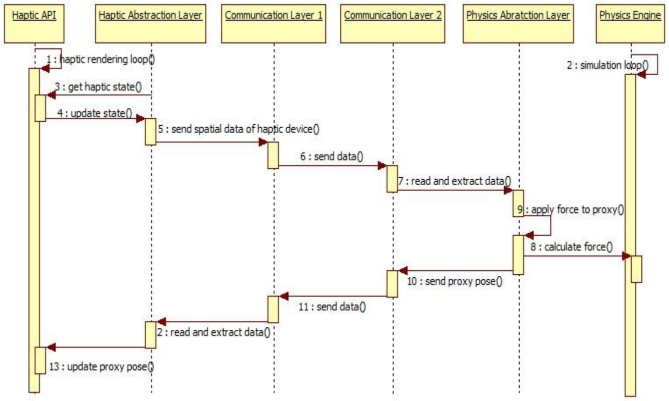


Figure 3.Sequence diagram of the framework.

B. Haptics In Physics Simulation

Two issues are concerned for stable haptic rendering in the integration of haptics in physics simulation. First, in most physics engines, dynamic objects are usually directly subject to applied forces instead of poses from external interactions. However, when dealing with a position-controlled impedance-style haptic device, the direct inputs from the user to the haptic device are poses rather than forces. That means the user cannot directly manipulate the objects in the simulation via pose control. It is therefore necessary to converts poses to forces.

The other issue is the disparity of data update rate as discussed in the previous sections. Only simple simulation applications are able to run at the haptic device's servo loop rate (typically about 1k Hz) while physics simulation is usually only optimized to run at the same rate as the graphics loop (30~60 Hz).

Virtual coupling technique [9] is commonly employed to achieve stable haptic rendering. In this technique, the simulation is stepped in a separate thread and a synchronization mechanism is used to update the positional inputs, as shown in Fig. 4. Each thread samples the respective spring-damper positions. The spring-damper used by the haptic device is attached to a proxy to update its position whenever the simulation is stepped. Similarly, the simulation samples the device position before each simulation step in order to calculate the input force that will be applied to the object being manipulated in the simulated environment.

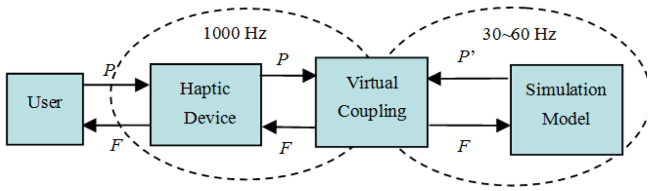


Figure 4.Virtual coupling method for stable haptic rendering.

Note that virtual coupling is usually applied in the same process. However, in our framework, the haptic rendering and physics simulation are executed as two separated processes running at different frequencies in an asynchronous manner. The server takes care of the major processes in the entire simulation, including collision detection and response, soft object deformation, and force calculation, whereas the client one only handles haptic rendering. Therefore, in this paper, the concept of virtual coupling is extended to an inter-process framework, namely remote virtual coupling, which will be explained in the next section.

C. Remote Virtual Coupling

Virtual coupling introduces an intermediate layer between the haptic device and the simulation. It is modeled by connecting a spring-damper system between the simulated body and the device end-effector. The spring-damper system provides a stable mechanism for the haptic device and the simulated body to exchange forces. Fig. 5 depicts the mechanism of the remote virtual coupling. In the client side, the haptic device is linked to a virtual object A with a spring-damper system. The pose of the virtual object is updated by the proxy in the simulation (server side) through the communication channels. The force is calculated at around 1k Hz based on the spring-damper model, whereas the pose of the virtual object is updated at 50 Hz. In the server side, the proxy is linked to a virtual object B based on the spring-damper model. The pose of object B is updated by the haptic device in the client side via the communication channel. The force calculated by the spring-damper model will be applied to the proxy, so that its pose will be updated at 30~60 Hz subject to the interactions between the remote haptic device and the local simulation environment. Different constants are used in the two spring-damper systems to compute the forces to be applied respectively to the device and the manipulated proxy in the simulation environment. The forces can be tuned to achieve a suitable trade-off between stability and smooth haptic feeling without evident sticking artifacts.

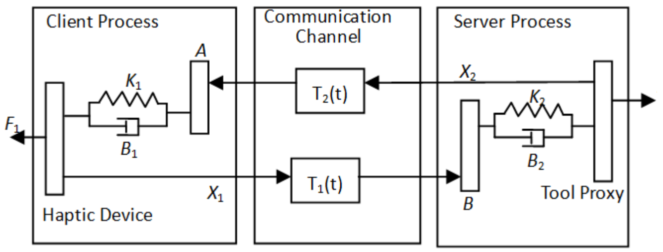


Figure 5.Schematic diagram of remote virtual coupling in IPC.

IV. BENCHMARK

The proposed remote virtual coupling technique is evaluated by comparing its performance with the interpolation-based haptic rendering method. In the interpolation method, the force is calculated at the server side based on the proxy manipulated by the haptic device and the end-effector pose of the haptic device using a spring-damper model. At the same time, the force is transferred from the server side to the haptic rendering loop in the client side by means of the IPC mechanism as mentioned in the previous sections. Since the force is calculated in server side at a graphic update rate (30~60 Hz) and the force needs to feed haptic rendering at a much higher update rate about 1k Hz, a force interpolation method thus needs to be adopted in order to resolve this disparity issue. Refer to Fig. 6, the haptic force is interpolated as,

$$f_j^h = \alpha[f_{i-1}^s + \frac{(f_i^s - f_{i-1}^s)}{N} \times (j\%N)] \tag{1}$$

$$N = T_h / T_s \tag{2}$$

$$f_i^s = K \times (P_i^p - P_i^h) - B \times V_i^p \tag{3}$$

where f_j^h is the interpolated force in the haptic loop at time step j within a time interval T_h, α is the scale factor, f_i^s is the calculated simulation force in the simulation loop at time step i within a time interval T_s as shown in Fig. 6, K is the stiffness constant, B is the damping coefficient, P_i^P is the position of the proxy, P_i^h is the position of the end-effector, and V_i^P is the velocity.

To compensate the delay due to update rate disparity between the physics simulation and haptic rendering process, an improved version of the equation (3) for calculating simulation force is developed by taking the predicted position of the proxy into account, i.e.,

$$f_i^s = K \times (P_{i+1}^p - P_i^h) - B \times V_i^p \tag{4}$$

where P_{i+1}^P is the predicted position of the proxy at time step, i+1

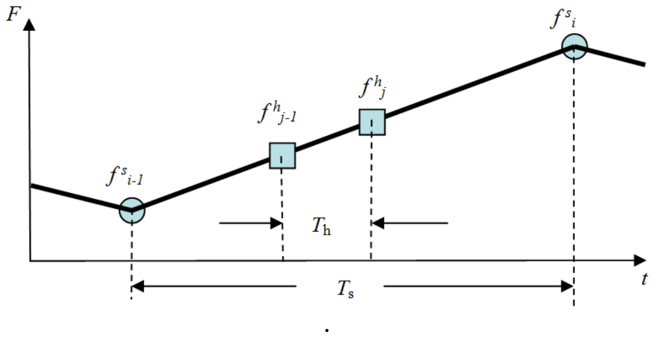


Figure 6.Force interpolation-based haptic rendering.

Several experiments are carried out to compare the performance of the remote virtual coupling technique with the interpolation-based haptic rendering algorithms. Here, the passivity and stability of two methods are compared. In the benchmarking experiments, the haptic device PHANTOM® DesktopTM from the Sensable Technologies is used. It has 6 degree-of-freedom positional input and 3 degree-of-freedom force output. The proximal end of the device is fixed with the secondary arm, as shown in Fig. 7.

The end-effector of the haptic interface is initially lift vertically upwards in the air at the position y=20 mm with zero initial velocity.

When the stylus is released, it drops in free fall due to gravity until it hits the virtual horizontal wall at the position, y= 0 within the horizontal X-Z plane. When it hits the virtual wall, the force calculated by a spring-damper model pushes the stylus vertically upwards. First, the proxy-based method is used, i.e. no special technique is used to streamline haptic rendering. Therefore, the trace of the end-effector in Y direction over time is shown in Fig. 8, which is in the form of a damping curve. It can be seen from the figure that the end-effector of the stylus bounces for about 0.7 second, and finally stays at the position, y=-2 mm.

The experiment is repeated using the two haptic rendering methods and the force-time curves are measured. With K=100 N/s, B=0.01 Ns/m and using equation (3) for force interpolation, it can be seen from Fig. 9 that the force oscillates for more than 4 seconds, which is too long and not favorable for rendering smooth haptic sensation. When the prediction of proxy position is considered, with equation (4), it is found that the force can quickly converge within 0.4 second, as shown in Fig. 10. When a higher stiffness constant is applied, say K=500 N/m, the force generated by the interpolation-based method diverges very quickly even though the prediction technique is used, while the force generated by remote virtual coupling can still converge as shown in Fig. 11. These experiments demonstrate that the remote virtual coupling technique outperforms the interpolation methods in terms of stability and robustness. The improved version of force interpolation method with proxy position prediction is a better approach than the conventional force interpolation method for stable haptic rendering.

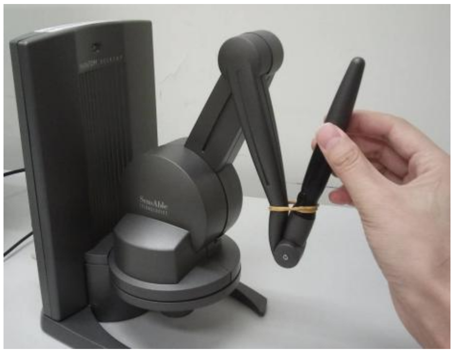


Figure 7.The haptic device, PHANTOM® DesktopTM used in the simulation system. The proximal end of the stylus is fixed with the secondary arm using a rubber band

V. PERFORMANCE EVALUATION

Two simulation scenarios, involving interactions of rigid and soft objects respectively, are implemented using the proposed framework to investigate the performance of the remote virtual coupling on the IPC platform and the interpolation-based haptic rendering algorithms. The experiments are performed on an Intel i7-2600 (3.4G Hz) personal computer running Windows 7, with 4 GB RAM, an NVIDIA Quadro 4000 graphics card, and a PHANTOM® Desktop as the haptic device.

A. Software Tools

Several software packages are used to implement the proposed haptic rendering framework, including Bullet (http://bulletphysics.org), Physics Abstraction Layer (PAL) (http://www.adrianboeing.com/pal), and Haptik (http://sirslab.dii.unisi.it/haptiklibrary/). Bullet is an open source physics engine which features 3D collision detection, soft body dynamics, fluid simulation, and rigid body dynamics. PAL is a high-level interface for low-level physics engines used in games, simulation systems, and other 3D applications. It also supports a number of dynamic simulation methodologies. Haptik is a component-based lightweight library that provides a hardware abstraction layer for access to haptic devices. Hardware from different manufactures can be easily accessed in a uniform way, enabling applications to be conveniently developed regardless of dependencies on specific APIs, hardware and drivers. The IPC is implemented with named pipes. A duplex pipe is employed for communication between the pipe server and one or more pipe clients. In Windows, the design of named pipes is based on client-server communication, working in a way like sockets.

B. Rigid Bodies

The dynamics of rigid objects is simulated by using the rigid body simulation package provided by the physics engine Bullet. The objects are manipulated with the haptic interface as shown in Fig. 12. In this experiment, the performance of the framework is evaluated by simulating 125, 250 and 500 rigid cubes in the virtual environment. The results in Fig. 13 show that the average update time per step is about 10 ms (100 Hz) for 125 cubes, 23 ms (about 43 Hz) for 250 cubes, and 57 ms (about 18 Hz) for 500 cubes. In the last case, the simulation update rate is less than 20 Hz, indicating that visual discontinuity can be perceived. As the simulation engine occupies a lot of computation, high update rate required for smooth haptic rendering cannot be guaranteed if haptics and physics simulation run on the same process.

When the multi-processing framework proposed in this paper is applied, haptic rendering and physics simulation are executed on separate processes. The computation resources of the former will not be deprived by that of the latter; even the physics engine may occupy a large amount of CPU time and memory resources. This is demonstrated with the experiments where the proxy is controlled by the haptic device to push through (along axis Z) a stack of 5×5×5, 10×5×5 and 10×10×5 cubes respectively, as illustrated in Fig. 14. The haptic and graphic update rates with regards to different simulation complexity are listed in Table 1. The haptic update rates are calculated based on the average update time in the haptic rendering loop. The sampling time lasts for about 20 seconds. It can be seen from the table that while the update rate of the simulation engine drops down from 100 Hz to 18 Hz as the number of rigid cubes increases from 125 to 500, the haptic update rate remains relatively stable around 1k Hz.

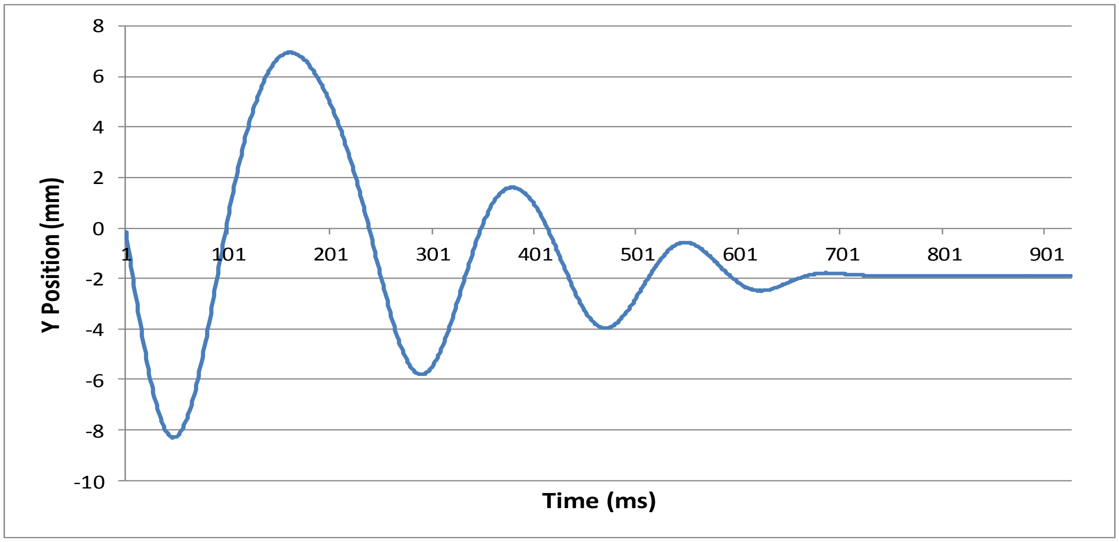


Figure 8.Position curves obtained with simple proxy-based method, which is executed in the haptic loop at 1k Hz) (K=200 N/m, B=0.02 Ns/m).

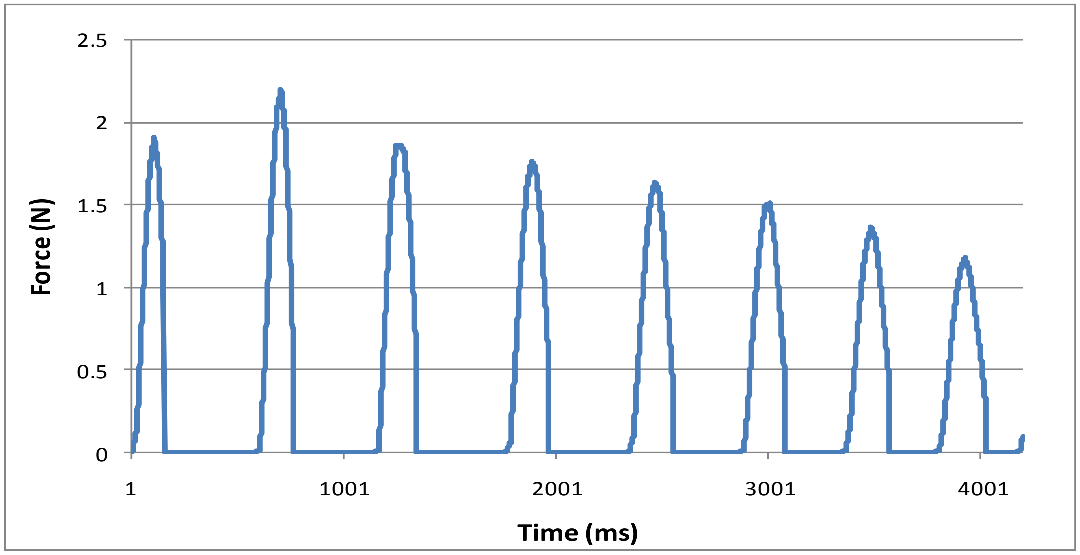


Figure 9.Force profile obtained using the interpolation-based method without prediction, with K=100 N/m, B=0.01 Ns/m

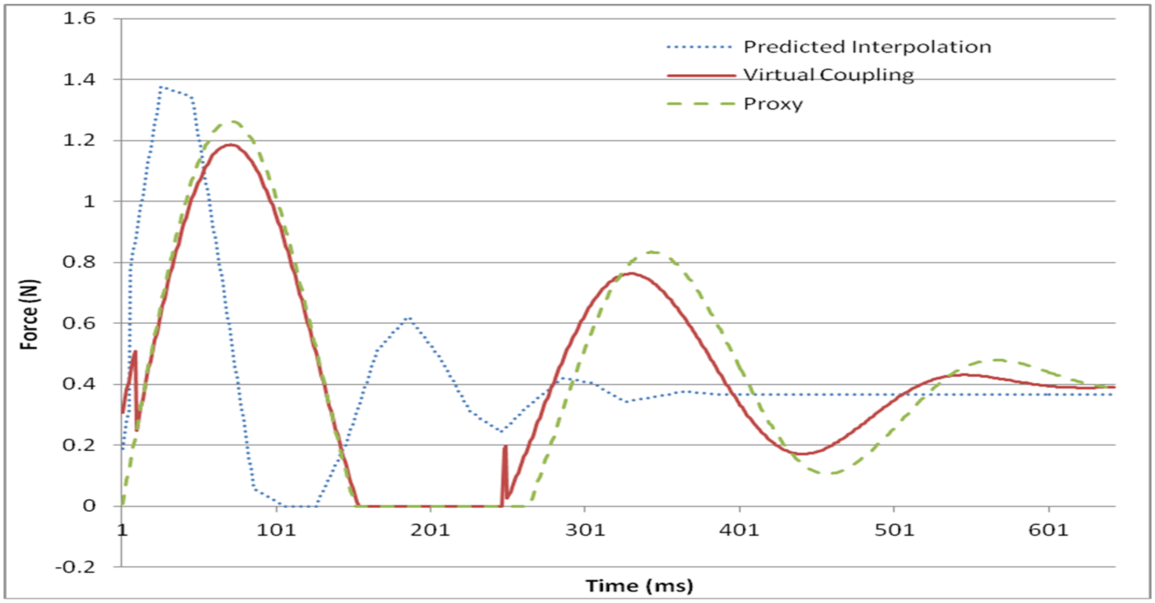


Figure 10.Force curves obtained with virtual coupling and the interpolation-based method with prediction (K=100 N/m, B=0.01 Ns/m). The green dotted line, referring to the force curve obtained by proxy-based haptic rendering at 1k Hz update rate, is provided as a reference

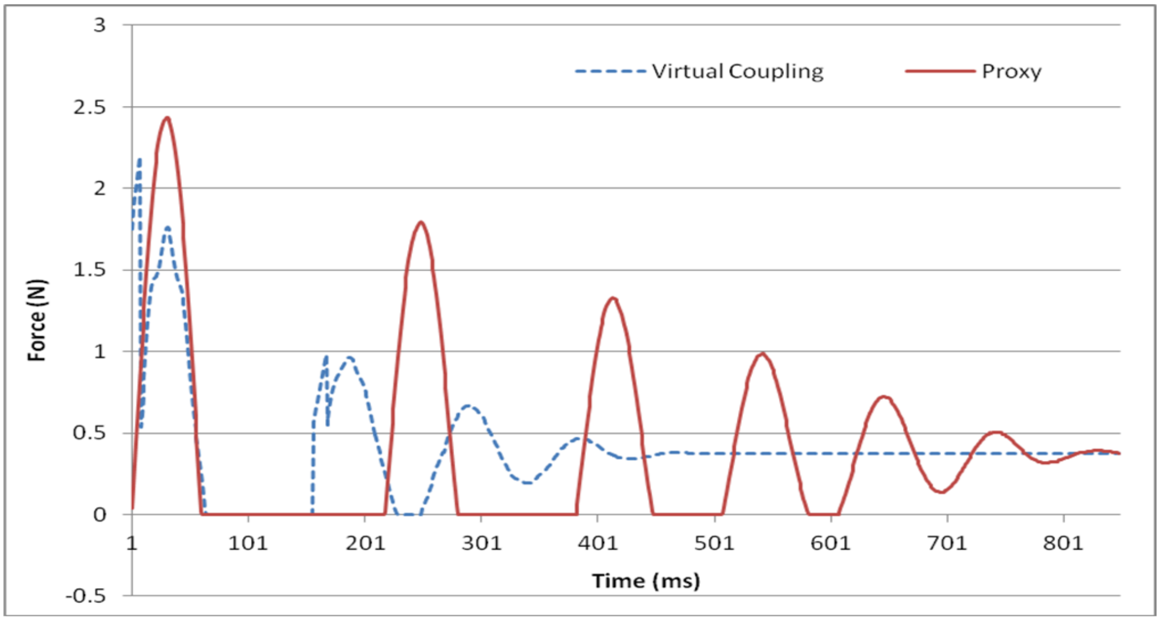


Figure 11.Force curves obtained using virtual coupling (K=500 N/m, B=0.05 Ns/m). The red solid line, referring to the force curve obtained by proxy-based haptic rendering at 1k Hz update rate, is provided as a reference.

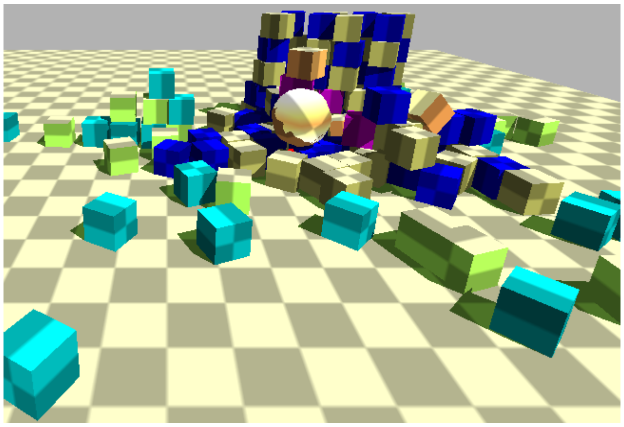


Figure 12.Screenshot of the simulated interactions of dynamic rigid cubes. The proxy (sphere) is manipulated by the haptic device to interact with the cubes.

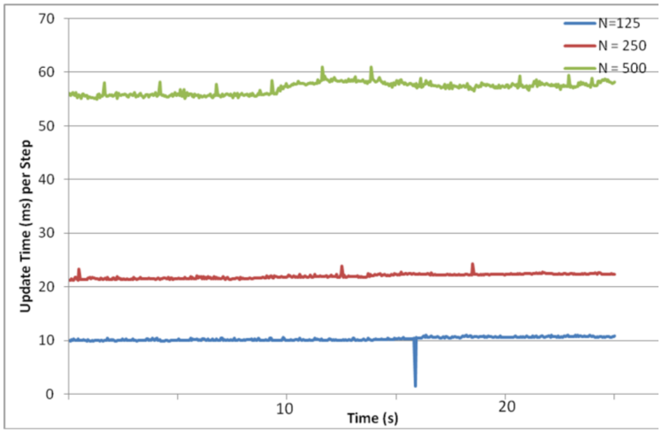


Figure 13.Time taken for one simulate step when the number of cubes N in the scene is 125, 250 and 500 respectively.

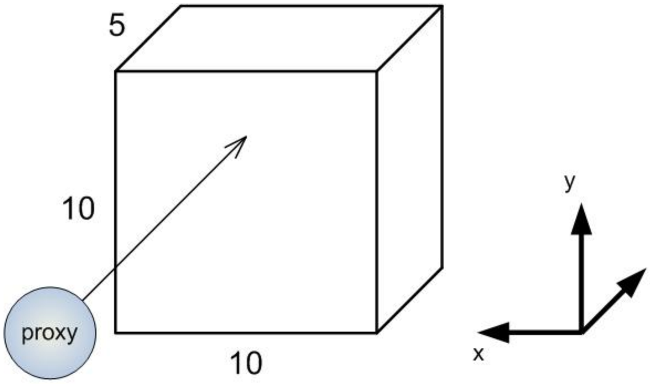


Figure 14.Pushing the proxy toward a stack of 10□10□5 cubes.

TABLE I. UPDATE RATES MEASURED FROM THREE SIMULATION SCENARIOS.

Number of Rigid Cubes	125	250	500
Simulation Engine Update Rate (Hz)	100	43	18
Haptic Rendering Update Rate (Hz)	1008	1003	1003

In the experiments, the corresponding pushing forces are recorded and plotted as shown in Fig. 15 and 16. Two different haptic rendering algorithms, force interpolation and remote virtual coupling are applied in the experiments. It can be seen from the figures that the forces calculated by the both algorithms are stable. While the force vibration can be perceived visually and kinesthetically in the interpolation-based haptic rendering algorithm, the forces generated by remote virtual coupling have relatively smoother profiles.

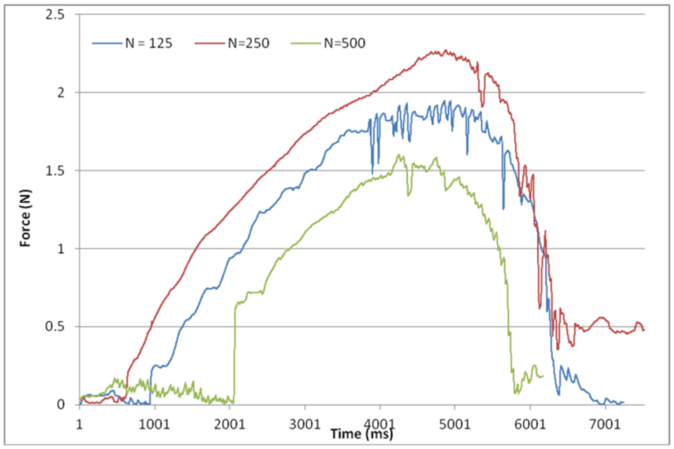


Figure 15.Force profiles measured when the proxy is moved toward a stack of 125, 250 and 500 cubes. The interpolation-based haptic rendering method is used in the experiment.

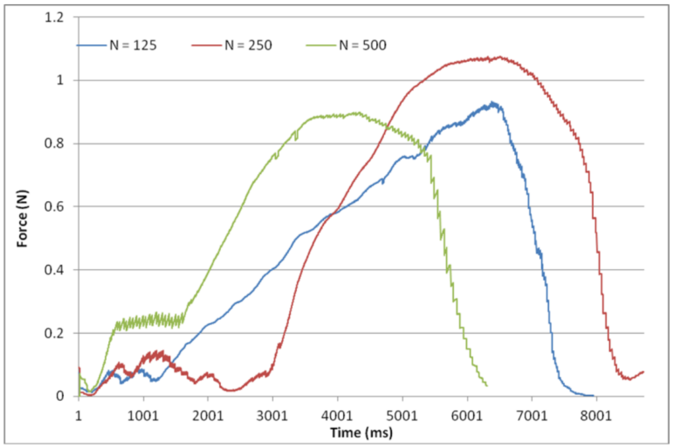


Figure 16.Force profiles measured when the proxy is moved toward a stack of 125, 250 and 500 cubes. The remote virtual coupling technique is used in the experiment.

C. Soft Objects

In this experiment, five soft bunny models are simulated in the virtual environment. As shown in Fig. 17. The bunnies can be touched by the proxy manipulated by the haptic device. Each bunny model has 450 nodes, 1353 edges and 902 faces. The contact forces generated by the interpolation-based method and the remote virtual coupling technique are plotted respectively in Fig. 18 and 19. It can be seen that the force profile of the remote virtual coupling is smoother than that of the interpolation-based method.

VI. DISCUSSIONS

While physics engines have been developed to reduce the effort required to simulate the physical world, they do not have much support for haptic rendering. In this study, research is conducted to facilitate the integration of haptic feedback into interactive applications developed with physics engines. To deal with the problem of update-rate disparity among the visualization, physics and haptics simulation processes, the proposed framework attempts to bridge the gap by making use of the techniques of remote virtual coupling technique and inter-process communication.

Notice that the purpose of the study is to propose a stable haptic rendering algorithm based on the framework rather than improving the computational performance or visual appearance of existing systems. Experiments have been carried out to evaluate the feasibility of the proposed method with the interpolation-based method. The results demonstrate that remote virtual coupling has better stability and robustness and that physics engines can also be used to support stable haptic rendering although this is not primarily designed for that purpose.

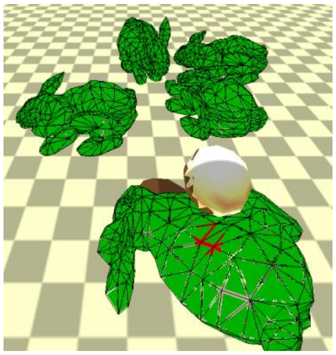


Figure 17.Screenshot of the interactions between the proxy and the soft bunnies.

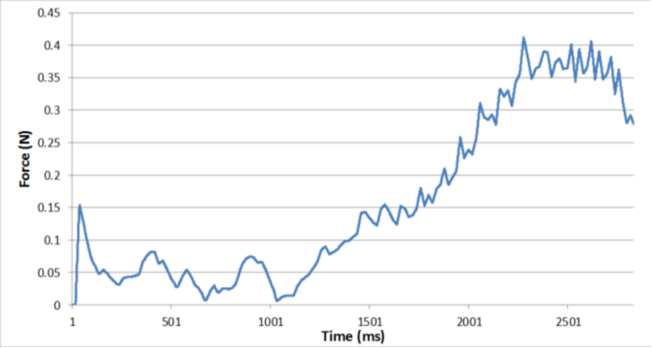


Figure 18Force profile recorded with interpolation-based haptic rendering.

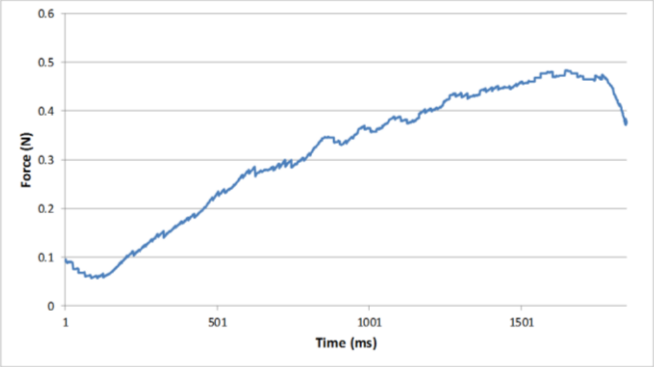


Figure 19.Force profile recorded with remote virtual coupling.

However, some latency and sticking force effects are observed even the parameters of the two spring-damper models, in the server and client side respectively, have been tuned carefully for remote virtual coupling in IPC. One possible solution is to use a massless proxy instead of the spring-damper method in server side. The poses of the proxy can be manipulated by the remote haptic device directly. However, objects in most physics engines are usually controll**e**d by forces and velocities rather than poses.

As illustrated in section V.B, the remote virtual coupling has relatively smoother force profiles in the cube-pushing experiment. It is worth noting that virtual coupling possibly filters out the useful vibration forces caused by the interactions between the proxy and the dynamic objects. This filtering effect is related to the setting of coupling parameters, i.e. stiffness and damping coefficients. Therefore, these parameters need to be tuned carefully based on the simulated body density, stiffness, size, friction and other physical parameters.

Although the framework based on the client-server architecture design can be possibly extended to networked haptics-enabled systems, it essentially runs on a single computer. To extend it for networked applications, it is necessary to deal with instability due to the stochastic nature of computer networks, e.g. time delay, jitter and packet loss, which are ignored in the communication protocol of the proposed framework.

VII. CONCLUSIONS

In advanced interactive applications, e.g. virtual surgery, haptic rendering plays an important role alongside graphics rendering and physics simulation. The use of physics engines is able to facilitate physically realistic simulation and improve the visualization of the simulation results, but it is yet to be fully compatible with haptic rendering.

The proposed framework fully decouples the haptic rendering from physics simulation by means of IPC and the remote virtual coupling techniques. It is demonstrated experimentally that this framework can guarantee stable haptic rendering even when the physics simulation runs at a low update rate. It also has the potential to reduce development time and effort by exploiting the benefits of physics engine.

Future work will be conducted to further demonstrate the flexibility of the proposed framework by employing other physics engines, e.g. PhysX (http://www.geforce.com), and hapitc devices, e.g. Novint Falcon (http://www.novint.com). Convenient integration of haptic rendering and physics simulation is anticipated since the abstraction layers for the respective processes shall provide convenient interfaces for switching the underlying physics engines and haptic interfaces.

ACKNOWLEDGMENT

This work was supported in part by the Research Grants Council of Hong Kong SAR (Project No. PolyU 5134/12E, 5152/09E) and the Hong Kong Polytechnic University (Project Account Code 87RF).

REFERENCES

[1] Basdogan, C., et al., Haptics in minimally invasive surgical simulation and training. Computer Graphics and Applications, IEEE, 2004. 24(2): p. 56-64.

[2] Lah, T., THUMP: an immersive haptic console for surgical simulation and training. Medicine Meets Virtual Reality 12: Building a Better You: The Next Tools for Medical Education, Diagnosis, and Care, 2004. 98: p. 272.

[3] Salisbury, K., F. Conti, and F. Barbagli, Haptic rendering: introductory concepts. Computer Graphics and Applications, IEEE, 2004. 24(2): p. 24-32.

[4] Otaduy, M.A. and M.C. Lin. Introduction to haptic rendering. in International Conference on Computer Graphics and Interactive Techniques: ACM SIGGRAPH 2005 Courses: Los Angeles, California. 2005.

[5] Maciel, A., et al., Using the PhysX engine for physics - based virtual surgery with force feedback. The International Journal of Medical Robotics and Computer Assisted Surgery, 2009. 5(3): p. 341-353.

[6] Love, L.J. and W.J. Book, Contact stability analysis of virtual walls. 1995.

[7] Choi, K.S., et al., Haptic Rendering in Interactive Applications Developed with Commodity Physics Engine. Journal of Multimedia, 2011. 6(2): p. 147-155.

[8] Salisbury, K., et al. Haptic rendering: Programming touch interaction with virtual objects. in Proceedings of the 1995 symposium on Interactive 3D graphics. 1995: ACM.

[9] Colgate, J.E., M.C. Stanley, and J.M. Brown. Issues in the haptic display of tool use. in Intelligent Robots and Systems 95.'Human Robot Interaction and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on. 1995: IEEE.

[10] Zilles, C.B. and J.K. Salisbury. A constraint-based god-object method for haptic display. in Intelligent Robots and Systems 95.'Human Robot Interaction and Cooperative Robots', Proceedings. 1995 IEEE/RSJ International Conference on. 1995: IEEE.

[11] Ruspini, D.C., K. Kolarov, and O. Khatib. The haptic display of complex graphical environments. in Proceedings of the 24th annual conference on Computer graphics and interactive techniques. 1997: ACM Press/Addison-Wesley Publishing Co.

[12] Susa, I., M. Sato, and S. Hasegawa. Multi-rate multi-range dynamic simulation for haptic interaction. in World Haptics Conference (WHC), 2011 IEEE. 2011: IEEE.

[13] Hasegawa, S., et al., Inter-process communication for force display of dynamic virtual world. ASME DYN SYST CONTROL DIV PUBL DSC., 1999. 67: p. 211-218.

[14] Mark, W.R., et al. Adding force feedback to graphics systems: Issues and solutions. in Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. 1996: ACM.

[15] Adachi, Y., T. Kumano, and K. Ogino. Intermediate representation for stiff virtual objects. in Virtual Reality Annual International Symposium, 1995. Proceedings. 1995: IEEE.

[16] Astley, O.R. and V. Hayward. Multirate haptic simulation achieved by coupling finite element meshes through norton equivalents. in Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on. 1998: IEEE.

[17] Cavusoglu, M.C. and F. Tendick. Multirate simulation for high fidelity haptic interaction with deformable objects in virtual environments. in Robotics and Automation, 2000. Proceedings. ICRA'00. IEEE International Conference on. 2000: IEEE.

[18] Duriez, C., C. Andriot, and A. Kheddar. A multi-threaded approach for deformable/rigid contacts with haptic feedback. in Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. HAPTICS'04. Proceedings. 12th International Symposium on. 2004: IEEE.

[19] Adams, R.J. and B. Hannaford, Stable haptic interaction with virtual environments. Robotics and Automation, IEEE Transactions on, 1999. 15(3): p. 465-474.

[20] Akahane, K., et al. A proposal of a high definition haptic rendering for stability and fidelity. in Artificial Reality and Telexistence--Workshops, 2006. ICAT'06. 16th International Conference on. 2006: IEEE.

[21] Otaduy, M.A. and M.C. Lin, A modular haptic rendering algorithm for stable and transparent 6-DOF manipulation. Robotics, IEEE Transactions on, 2006. 22(4): p. 751-762.

[22] Ortega, M., S. Redon, and S. Coquillart. A six degree-of-freedom god-object method for haptic display of rigid bodies. in Virtual Reality Conference, 2006. 2006: IEEE.

[23] Glondu, L., M. Marchal, and G. Dumont, A new coupling scheme for haptic rendering of rigid bodies interactions based on a haptic sub-world using a contact graph. Haptics: Generating and Perceiving Tangible Sensations, 2010: p. 51-56.

[24] Yasrebi, N. and D. Constantinescu. Wave filter bank for high fidelity passive multirate haptic interaction with slowly updated virtual environments. in Haptics Symposium (HAPTICS), 2012 IEEE.

AUTHORS PROFILE

Xue-Jian He received his Ph.D. degree in mechanical engineering from the University of Hong Kong. He is currently a research associate at the School of Nursing, the Hong Kong Polytechnic University. His research interests include haptic modeling, human-computer interaction, medical simulation, virtual reality in medicine and health care.

Kup-Sze Choi received his Ph.D. degree in computer science and engineering from the Chinese University of Hong Kong. He is currently an assistant professor at the School of Nursing, the Hong Kong Polytechnic University. His research interests include computer graphics, virtual reality, physically based simulation, computational intelligence, and their applications in medicine and health care.

FF-MAC : Fast Forward IEEE 802.15.4 MAC Protocol for Real-Time Data Transmission

Khalid EL GHOLAMI
dept. of physics
STIC, Chouaib Doukkali University
El Jadida, MOROCCO dept. of computer science
LIMOS UMR 6158 CNRS, Blaise Pascal University Clermont-Ferrand, France

Najib ELKAMOUN
dept. of physics
Chouaib Doukkali University
El Jadida, MOROCCO

Kun Mean HOU
dept. of computer science
LIMOS UMR 6158 CNRS, Blaise Pascal University
Clermont-Ferrand, France

Abstract—**This paper presents a Fast Forward MAC layer designed for hard real-time applications in wireless sensor networks. This protocol is an enhancement to the IEEE 802.15.4 standard MAC layer proposed for Low-Rate Personal Area Network. The energy conservation mechanism proposed by the current standard is quite efficient and very flexible. This flexibility comes from the ability to configure different duty cycles to meet specific application's requirements. However, this mechanism has a considerable impact on the end-to-end delay. Our approach resolves the energy delay trade-off by avoiding the storage of the real-time data in the coordinator during sleep time. A new superframe structure is adopted and a deterministic reception scheduling is used. All the simulations were done using the network simulator 2 'NS-2'. The simulations outcomes show that this new proposed protocol performs better than the current standard and reduces considerably the end-to-end delay even in low duty cycle networks. Our protocol can also provide a delay bound for all network configurations which allows a better choice of the duty cycle for the required delay.**

Keywords-component; IEEE 802.15.4; WSN; Superframe; star topology; delay; Duty cycle; D-GTS

I. Introduction

Recent advances in Microelectronic Mechanical Systems (MEMS) and wireless communication technologies have made wireless sensor network or Internet of Things (IoT) one of the most important research fields during the last years. This type of network is distinguished from other wireless ad hoc networks by its unique characteristics; namely, limited memory and processing power, high energy constraint, high node density and hardly unreliable (lossy) wireless communication. These constraints are challenging and open many research perspectives in different areas of interest. The first research works were interested on increasing the node life-time by minimizing the power consumption. Since this energy is mainly consumed by the radio transceiver, many works were led to resolve this problem by enhancing the communication protocols in different OSI model layers 'Open System Interconnection'. In our work, we focus on the enhancement of the medium access control 'MAC' sub-layer to minimize the end-to-end delay for time sensitive applications when considering a low duty cycle. We proposed a new IEEE 802.15.4-like MAC protocol that enhance the GTS (Guaranteed Time Slot) mechanism provided by the IEEE 802.15.4 standard [1] and bypass its limitations identified in this paper. Our work is based on the enhanced superframe structure of the IEEE 802.15.4 standard proposed in [2] (hereinafter, we will refer to this work as 'enhanced superframe'). This new superframe structure gives the time sensitive packets the possibility to be sent and received in the same superframe and, by consequence, minimizes the probability of storing them during sleep period in the coordinator queue. In this paper we identified some limitation of this proposal and we provide an important enhancement to this new superframe structure by providing a deterministic medium access algorithm in reception mode to avoid the randomness introduced by CSMA-CA algorithm and to ensure the acquisition of timely information from source to destination.

A. General problem Description

The usage of a low duty cycle allows the network nodes to save the battery power by switching on and off the radio alternatively. According to the standard, typical applications for IEEE 802.15.4 devices are anticipated to run using a very low duty cycles (under 1%); this duty cycle is translated to a long sleep time (see equations (4) and (5)). In star networks or when some network nodes use a GTS to send their critical data; all packets have to be sent first to the coordinator which is responsible for forwarding them to their final destinations. When the packets are received by the coordinator, they are stored in its queue until the next superframe. Then the destination node can pull the pending data after a reception of the beacon frame. This process forces the coordinator to store packets during sleep time. Moreover, in this type of scenarios, the node may remain inactive for a long time which increases the communication latency, since during sleep time; data may have to wait until the next active portion (CAP 'Contention Access Period') located in the next superframe to start the transmission.

B. Our contributions

We showed in [2], that the enhanced superframe structure outperforms the standard and resolve partially the energy-delay tradeoff. The new MAC protocol proposed in this paper provides a deterministic medium access, and delay bound for Hard Real-Time 'HTR' applications. This protocol is unaware of the sleep time length since data can reach its destination before the inactive portion of the superframe. Accordingly, the end devices using our protocol may send and receive critical data in short time and go to sleep to save power. These two works are discussed later in more detail.

The rest of this paper is organized as follows. Section 2, review the IEEE 802.15.4 standard. In section 3, we discuss some related works. While in section 4, we identify some week point of the enhanced superframe. Then we explain our new proposed protocol. In section 5, we show the performance evaluation study of the new protocol compared to the standard and the 'enhanced superframe'. We finish this paper by a conclusion and some perspectives for the future works.

II. Overview of the IEEE 802.15.4 MAC layer

The IEEE 802.15.4 standard is one of the main communication protocol designed to meet the requirement of the wireless sensors networks and IoT. This standard specifies the medium access control sublayer 'MAC' and Physical layer 'PHY' for low rate Wireless Personal Area Networks 'LR-WPAN'. In this section we will focus only on the MAC sublayer and its different parameters, since it's the subject of our contributions.

An IEEE 802.15.4 node can operate in two alternative modes: (1) the beaconless mode, where the nodes use only the unslotted CSMA-CA protocol to randomly manage channel access and avoid collisions. Since the IEEE 802.15.4 frame size is very small, this modified version of the standard CSMA-CA algorithm doesn't use the RTS/CTS mechanism to resolve the hidden terminal problem. The synchronization is not needed and the Quality of Service 'QoS' mechanisms are not provided in this mode, which makes it more suitable for applications without QoS requirements. And (2) the beacon-enabled mode, that uses a superframe to control the channel access. The superframe structure may be divided into three periods (see Figure 1): (1) contention access period (CAP), where network nodes use the slotted version of the CSMA-CA algorithm to contend for the channel access. (2) Collision free period (CFP) where the channel is reserved and can be used exclusively by the reserving node using a slot labeled Guaranteed Time Slot 'GTS'. The CFP period is optional and used by low-latency applications or applications requiring specific data bandwidth. And (3) the inactive portion (sleep period), which is also optional and used when the network nodes don't need to be awake all the time (suitable for most of wireless sensor network applications to minimize energy consumption).

In beacon enabled mode, the entire PAN 'Personal Area Network' is managed by the PAN Coordinator. It advertises periodically a packet named 'beacon' at the beginning of the superframe. This beacon is used to synchronize the attached devices, to identify the PAN, and to describe the structure of the superframe. It may also provide additional information about the pending addresses and the GTS configuration if needed. The superframe periods timing relay on the following parameters: beacon order (BO), the superframe order (SO) and the Final CAP Slot, where $0 \leq SO \leq BO \leq 14$. These parameters are specified in the beacon superframe which allows the network nodes to determine the superframe structure (the active period, The Contention Access Period 'CAP' length, the sleep time duration and the slot duration). The formula (1), (2) and (3) are used to calculate these durations:

$$BI = aBaseSuperframeDuration \times 2^{BO} \text{ (symbols)} \quad (1)$$

$$SD = aBaseSuperframeDuration \times 2^{SO} \text{ (symbols)} \quad (2)$$

$$sd = aBaseSlotDuration \times 2^{SO} = SD/16 \text{ (symbols)} \quad (3)$$

Where aBaseSuperframeDuration and aBaseSlotDuration are two constants predefined by the standard as 960 and 60 symbols respectively and denote the minimum length of the superframe and the slot respectively. Each symbol corresponds to 4 bits. BI (beacon interval) is the length of the whole superframe (including active period and inactive period). It is bounded by two beacon transmissions. The SD (superframe duration) represents the active period duration. And the 'sd' (slot duration) is the sixteenth of the active period.

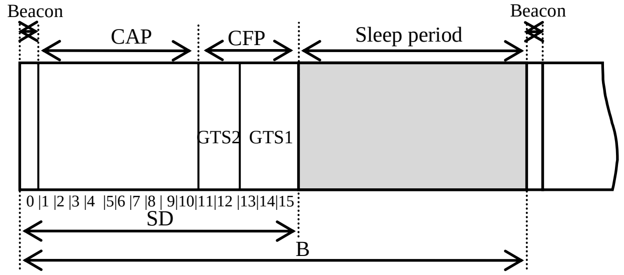


Figure 1 : IEEE 802.15.4 superframe structure

In the beacon-enabled mode, the PAN coordinator may allow the other network nodes to reserve a dedicated time slots to satisfy the bandwidth and latency requirements via a TDMA-like 'Time Division Multi Access' medium access method. These slots are labeled as GTS. Each node can allocate up to two GTSs (one for receive and one for transmit), and one GTS may have more than one slot. The number of GTSs is at most seven. These contiguous time slots form a Contention Free Period (CFP) which is placed at the end of the active period of the superframe. To use the GTS, the node has to send a GTS request to the PAN coordinator in the CAP (Contention Access Period), and when this request is honored, the coordinator will advertise in its beacon all the information related to the GTS allocation. The node has to keep tracking the beacon for any possible changes (deallocation or reallocation). If the node does not receive the beacon, it is not allowed to use its GTS and has to wait for the next beacon. The transmission during the GTS is indirect (i.e. data has to go through the coordinator, and then the coordinator advertises the pending address in the beacon so that the destination can poll it by sending a data request MAC command).

The energy limitation in WSNs is one of the most challenging aspects involved when designing protocols and considering QoS support in the network. This energy is directly related to the lifetime of the network. As we mentioned in the previous section, the IEEE 802.15.4 provides also a mechanism for power saving. This feature is possible only in beacon-enables mode when the BO is different than the SO (SO<BO).

This sleep-awake scheme is suitable for wireless sensor networks since the nodes do not need to stay awake all the time, they may operate for a short time to send or receive collected data. This mechanism allows the devices to save power during sleep time.

However, the choice of a low duty cycle is made at the cost of a higher latency. Since during sleep time, data may have to wait until the active portion of the next superframe to start the transmission. This time can be computed as the ratio between the superframe duration and the beacon interval that can be related to BO and SO via the following equation:

$$DC = SD/BI = 2^{SO-BO} \qquad (4)$$

$$\text{Sleep time} = 2^{BO} - 2^{SO} \qquad (5)$$

III. Related works and background

One of the most difficult problems to resolve in wireless sensors networks is the energy-delay tradeoff. The first MAC layers proposed in this field tends to reduce the power consumption since energy is a critical resource in wireless sensor nodes. For instance, S-MAC [3], T-MAC [4] H-MAC [5], X-MAC [6], WiseMAC [7], U-MAC [8], M-cube [9], RMAC [10] and Z-MAC [11] are duty cycle based MAC protocols that can specify sleep and wake up times for network nodes within the frame. The IEEE 802.15.4 standard can also be configured to operate in this mode.

However, in recent years, many WSN and IoT applications appeared and many of them require a certain level of QoS ‘Quality of Service’ for time sensitive data. In some applications, the information transported in the network may lose its meaning or may have a negative effect when it reaches the destination too late. Hence, QoS may be as important as the energy conservation in these applications. For instance, this fact leads the ISA100 group that standardizes wireless systems for industrial automation application, to specify in the ISA100.11a standard [12] different level of quality of service (classes from 0 to 5), depending on the importance of message timeliness. Reference [13] made a survey on real-time QoS support in wireless sensor networks and presented some real-time solutions including MAC and routing protocols, data processing strategies and cross-layer designs.

As we presented in the previous section, the IEEE 802.15.4 standard has proposed the GTS mechanism to meet these requirements. However, the standard presents some limitation identified by many researcher and many works were led to improve the GTS mechanism proposed by this standard. These works were interested on different aspect such as GTS allocation, GTS management and GTS efficiency. In [14], the superframe were extended to increase the number of GTS. The aim is to reduce the waste of channel bandwidth and to enhance the QoS support for multiple devices. Reference [15] divided the GTS length to slots smaller than a standard superframe slot to minimize the waste of the channel bandwidth. In [16] an implicit GTS allocation mechanism (i-GAME) is proposed. His protocol uses the round-robin algorithm to share the GTS by several nodes. [17] Proposes a method to resolve the insufficient GTS slot problem (that are limited to 7 by the standard specification) by allocating the GTS with higher priority first. The GTS requests are classified according to their priorities which allow GTSs to be allocated first for nodes having real-time data by giving them higher priorities. It overcomes the under utilization of GTS bandwidth and the number of the concurrently allocable GTSs. [18] proposed a fully deterministic MAC protocol that supports a predefined time slots used for real-time association. This new scheme tends to avoid unsuccessful GTS request and to avoid also the collision during the GTS between nodes of different clusters in the same transmission range. Other works [19][20][21][22][23][24][25][26][27] were interested on the improvement of the CSMA-CA algorithm to add QoS support for real-time applications.

In [2], we identified some other limitations related to energy-delay tradeoff and we proposed a new superframe structure to allow sending and receiving the real-time packets in the same superframe to avoid storing data during sleep time that may be very long. The simulation results proved that this new protocol has decreased the end-to-end delay compared to the current standard even when the network uses a low duty cycle. This protocol is covered in more detail in the next section since we propose in this paper a solution for its limitations that we identified.

Most of these algorithms and improvements can be easily adapted to our new protocol to improve and optimize the GTS usage.

IV. Proposed Protocol

A. *Enhanced superframe structure*

We proposed in [2] an enhanced superframe structure of the current IEEE 802.15.4 standard. This new superframe structure allows a faster access to the channel and avoids a high additional delay caused by the sleep time for time sensitive data. This proposal tends to minimize the end-to-end delay, even when considering a very low duty cycle, by sending and receiving the real-time data in the same superframe.

This new superframe has the same periods defined by the IEEE 802.15.4 standard (i.e. contention access period, contention free period and sleep time). The beacon is also sent at the beginning of each superframe and contains all information about it. However, in the new superframe, the CFP is placed after the beacon transmission. The CAP is placed between the end of CFP and the end of the active portion. This new scheme is very important and gives three main improvements. (1) Nodes with real-time data can access the channel faster than those having normal data, since they don’t need to wait for the end of the CAP to send their data. (2) The real-time nodes don’t need to contend for the channel access in the CAP, since they send all their data in the CFP period which is placed at the beginning of the superframe. This new scheme may improve the performance of the other nodes and decrease the bandwidth and energy wastage due to unnecessary contentions. (3) The third improvement is very important since it is related to the energy-delay tradeoff. This protocol gives the possibility to the real-time data to be sent and received in the same superframe. Hence, we avoid the additional delay caused by storing data in the coordinator during sleep time. After the end of the CFP period, the coordinator need to inform the network nodes about the new packets sent in the previous CFP period. For this purpose we created a new packet labeled Pending Real-Time Packets Advertisement ‘PRTPA’ that contains a list of all destination nodes having pending real-time data. Thereby, these nodes will send a data request command to the coordinator to poll this data in the CAP of the same superframe.

This new superframe has shown an important enhancement on the end-to-end delay since the data storage depends less on the sleep time. However, the use of the CSMA-CA algorithm even with an enhanced version made the delay to be dependent on the number of nodes. The randomness of the CSMA-CA method allows other non real-time packets to gain channel access before real-time ones. Hence, sometimes delay sensitive data may be stored during sleep time. Moreover, the usage of this mechanism for this data introduce energy and bandwidth wastage caused by the CSMA-CA (backoffs and randomness)

Hence, the usage of a weighted version of the CSMA-CA algorithm in this approach represents its main week points. The different priorities (Real-time and Best-effort data) were translated into weighted backoffs by using a shorter backoff interval for Real-time data than the Best-effort one. The simulations show that this method improved the delay performance compared to the current standard.

However the measured delay still depends on the nodes density because the used CSMA-CA does not insure the channel access to the Real-Time data. It only increases the successful channel access probability without any guarantee. This may force the coordinator to keep some critical data packets during the sleep time. This problem is illustrated in Figure 2 that shows how the end-to-end delay increases when the density increases.

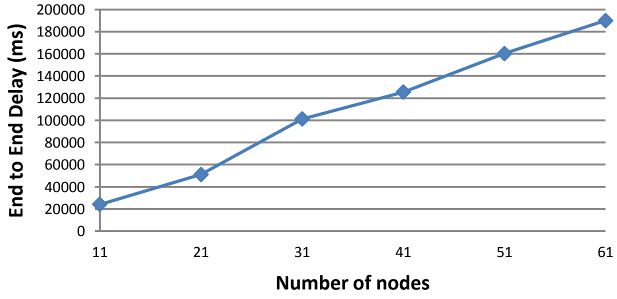


Figure 2 : End-to-End delay for different network sizes using the enhanced superframe [2]

To solve all these problems, we proposed a new protocol named ‘FF-MAC’

B. FF-MAC

FF-MAC stands for Fast Forward MAC protocol. This new protocol is designed to allow data transmission inside the cluster in a very short time. This new MAC protocol is based on the IEEE 802.15.4 standard by using the enhanced superframe structure presented above.

In our protocol, we propose the use of a deterministic medium access schedule to receive HRT data during CAP. In FF-MAC we keep the same enhanced superframe structure, the changes affect only the CAP which will be separated into two periods as described in Figure 3: (1) CAP for normal data and MAC commands packets. The network nodes will use in this period the standard CSMA-CA to send and receive. (2) The D-CPF, which is a new period dynamically created by the coordinator. The coordinator uses this period to send HRT data to their corresponding destination nodes in a contention free way. Since the coordinator has a clear view about the QoS requirements after receiving real-time data during CFP, it can select the destination nodes concerned by the real-time pending data and create a TDMA schedule forming a D-CFP period, the algorithm 1 is used in this case. Same as the CFP, The D-CFP period is formed by a set of contiguous D-GTS ‘Dynamically allocated GTS’. All these GTSs have to be set to receive only mode. This period appears only when there is some pending time sensitive data in the coordinator pending packet queue.

For efficiency purposes, these dynamically reserved GTSs are not related to the number of slots, but to the number of pending packets. Hence, we avoid a reservation of periods longer than what is needed. The coordinator calculates the required duration for each destination and creates a TDMA schedule. This schedule is sent to the network nodes using the new PRTPA packet; the message sequence chart is described in Figure 4.

This time depends on the packet size, number of pending packets, the bandwidth provided by the PHY layer, time needed for Acknowledgement packet if required and the IFS needed. The Pull data request MAC command is not needed to retrieve the pending data from the coordinator. The destination node only needs to switch there transceiver to receive mode (see algorithm 2).

ALGORITHM 1: FOR THE COORDINATOR

```
- Sending the beacon at the beginning of the superframe
  (including the pending data)
- Exchanging real-time data in the CFP period.
- If the coordinator has received data in the CFP
    { Send ”PRTPA” packet with D-CFP schedule
- Else
    { Send empty “PRTPA” packet to trigger the start of the CAP.
```

However, since the CAP length is limited and some packets need to be sent during CAP (e.g. management packets), the PAN coordinator shall preserve the minimum CAP length of aMinCAPLength and take preventative action if this value is not satisfied.

The CAP minimum length ‘aMinCAPLength’ is the same as the one defined by the IEEE 802.15.4 standard as time needed for 440 symbols. If the D-CFP reaches the maximum limit, the coordinator will stop the process and reserve D-GTS only for the first nodes.

ALGORITHM 2: FOR END-DEVICES

```
- If "PRTPA" is received
  {
    - If "PRTPA" packet has pending data
      {
        - If "PRTPA" packet advertise current node address
            { Delay until the corresponding D-GTS and change
              the transceiver status to RX_ON (reception mode)
        - Else
            { Go to sleep until the end of the D-CFP period.
      }
    - Else
        { Start CAP as described in the IEEE 802.15.4 standard
  }
- Else  // "PRTPA" is lost
    { Delay until the (end of CAP – aMinCAPLength) and
      start the transmission
```

As we noticed earlier, these new GTSs are dynamic which means that they appear only when needed and if no Real-Time data is pending; the corresponding D-GTS will disappear immediately. The D-GTS size is different than the corresponding GTS size, since it depends on the number of packets and not the standard slot size.

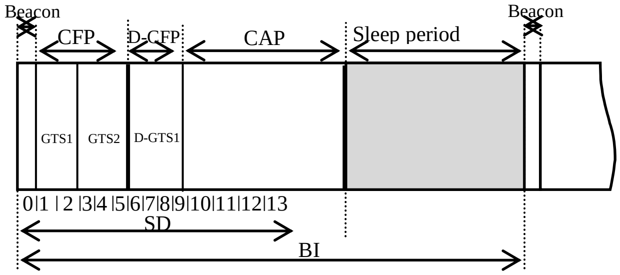


Figure 3 : Superframe structure according to FF-MAC protocol

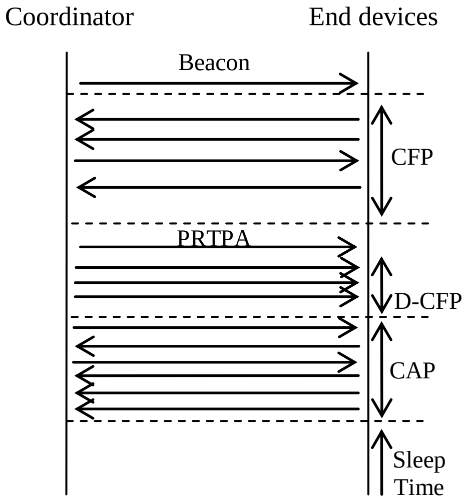


Figure 4 : Message sequence chart describing the packet exchange timing

Since the WSN communication is unreliable, the network nodes may miss some important packets. One of these packets is the update packet (PRTPA) that we proposed. Hence, if it's lost by a node, the latter will miss the information about the new D-CFP (which is dynamically changed in every superframe, depending on the pending real-time packets); For this reason, this node will suppose the worst case where the CAP reaches the minimum duration and will delay its data transmission until the last aMinCAPLength time before the end of the CAP. Then, it will try to send its packets (see algorithm 2). This handling will allow the protection of D-CFP period from unexpected collisions due to PRTPA packet loss.

V. PERFORMANCE EVALUATION

The performance evaluation simulations were built upon ns-2 [28] 'network simulator 2' (version 2.34) using the WPAN 'Wireless Personal Area Network' model [29] that simulate the IEEE 802.15.4 standard. The GTS management is missing in the official ns-2 version. It was implemented in our previous work [2].

The simulations make the following assumptions. For the physical layer, we use the IEEE802.15.4 PHY 2.4 GHz that provides 250 kbps. The IEEE802.15.4 MAC layer operates in beacon enabled mode since the GTS mechanism is only allowed in this mode. We use a star topology in all our simulations. All the scenarios are similar and contain the PAN coordinator which is placed in the center of the star network to reach all the network nodes, and a variable number of nodes randomly distributed over a 15m radius circle. The routing protocol is disabled since we evaluate our approach without the influence of the upper layers. In our simulation we disabled ARP (Address Resolution Protocol) since it's not needed in ZigBee networks. For MAC layer reliability, all the packets require MAC layer acknowledgement. The application layer uses 50 bytes UDP packets with data rate of one packet each BI (Beacon Interval) since in real word, the BO may be chosen depending on the sensing frequency. The traffic load is set by varying the number of network nodes.

In this performance evaluation section, we compare three protocols; The IEEE 802.15.4 standard, the enhanced superframe, and the FF-MAC proposed in this paper.

Many important and common points are highlighted by the simulation results. First, all these scenarios show that the end-to-end delay is considerably minimized using FF-MAC in all scenarios. Second, the sleep time has no impact on this delay when using our new protocol. Third, the node density does not influence the delay for real-time data. We will discuss in the following these results in more detail.

Figure 5 show the evolution of the end-to-end delay for different beacon order values, SO 'Superframe Order' is fixed to 5. These results are obtained for a 21 nodes network. We can clearly notice that the enhanced superframe approach provides a better delay performance than the current IEEE 802.15.4 standard. However, the delay increases considerably for both of them when the sleep time increases. As we discussed in section 4, both of these protocols use the random CSMA-CA algorithm to receive time sensitive data. The randomness inherent to this algorithm may force the coordinator to store data during the sleep time. The difference between these two algorithms is that in the enhanced superframe we have the possibility to send and receive data in the same superframe, and we used a weighted version of the CSMA-CA which gives a higher probability to access the channel for emergency data among normal data. In the same figure we notice that our FF-MAC protocol resolved this problem and allows a very fast transmission that doesn't depends on the sleep time. This is exactly what was expected by our proposed protocol since we replaced the random CSMA-CA algorithm by a deterministic scheduling. The delay is very low and we can provide a delay bound for different scenarios which is not possible in the IEEE 802.15.4 standard.

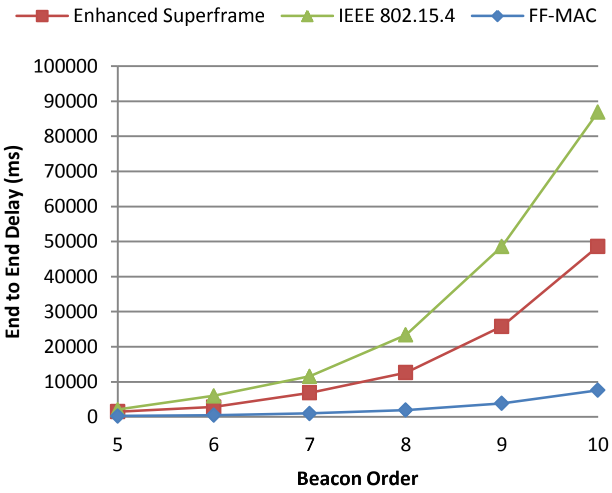


Figure 5 : End-to-End delay vs. beacon order (21 nodes, SO=5)

In Figure 6, we increased the nodes density in the network to evaluate its impact on the end-to-end delay, we can notice that the delay has increased considerably for both of the first protocols (enhanced superframe and IEEE 802.15.4) while the FF-MAC provide nearly the same delay shown in the Figure 5. We can explain this by the usage of the D-CFP which allows a dynamic reservation of the bandwidth for packets with delay constraint. These packets are sent without contention.

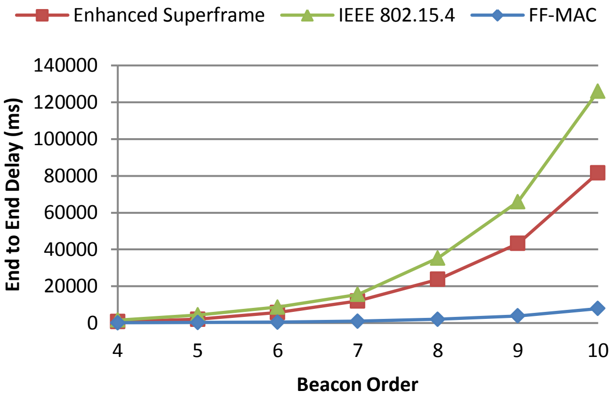


Figure 6 : End-to-End delay vs Beacon Order for (N=31, SO=4)

Figure 7, shows the behavior of the measured delay for various duty cycles. The results prove for all operation modes that our approach provides a very short delay if compared to the other protocols.

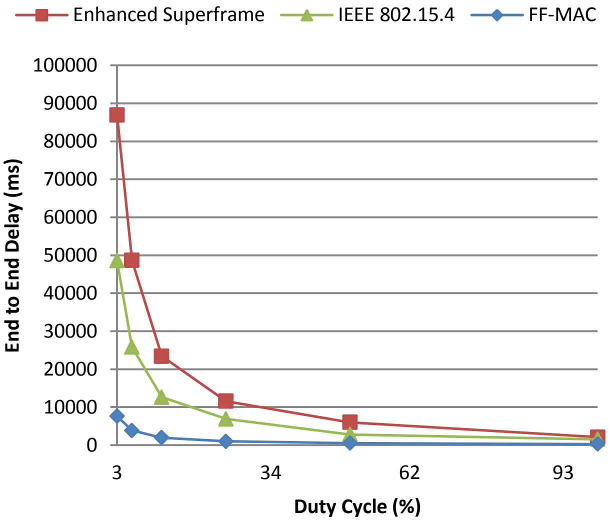


Figure 7 : End-to-End delay vs. duty cycle (21 nodes)

In Figure 8, we measured the end-to-end delay against the beacon order for different node numbers to evaluate the behavi or of FF-MAC when the number of the network nodes increases. For all chosen densities (21, 31, 41, 51, 61, 71 and 81), the FF-MAC provides a very stable delay that doesn't follow the density changes. Our deterministic algorithm provides the required quality of service even in a dense network. The impact of the node density on the delay is also shown in Figure 9, where BO and SO are fixed to set a very low duty cycle (BO=10, SO=4: DC = 1.56%). We can easily notice that the density has no impact on our FF-MAC, while it increases considerably the delay for the other protocols.

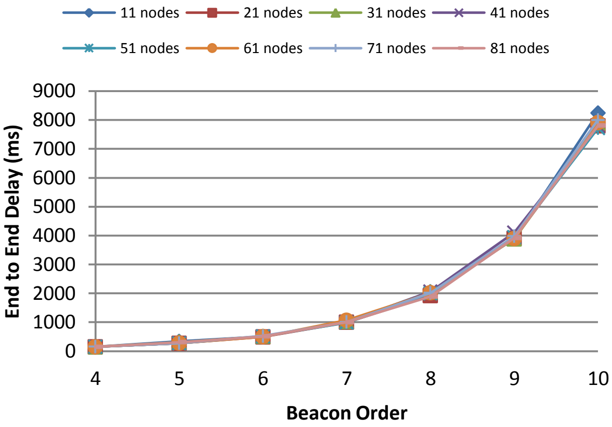


Figure 8 : End-to-End delay vs. BO for different nodes densities

All these presented results prove the enhancements expected by our approach. Figure 10 show a summary of all simulation results using the protocol proposed in this paper. FF-MAC provides better performance in all these scenarios.

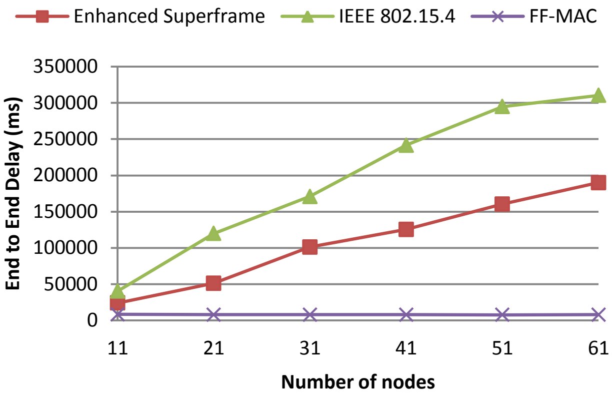


Figure 9 : End-to-End delay for different network sizes

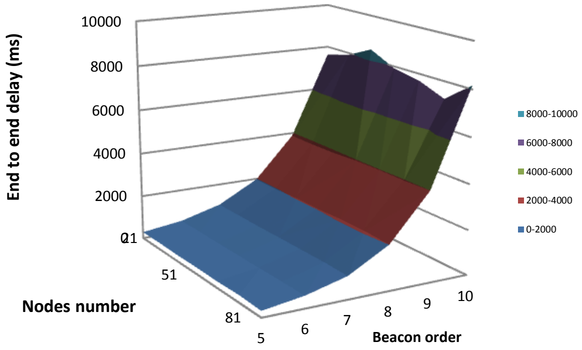


Figure 10 : Summary of all simulation scenarios of FF-MAC, SO = 4

VI. Conclusion And Perspectives

In this paper we made an overview of the IEEE 802.15.4 MAC layer and the enhanced superframe structure of this standard. We presented some limitation of the new superframe structure that is mainly related to the usage of the CSMA-CA algorithm. Then we presented a new IEEE 802.15.4-like MAC protocol named ‘FF-MAC’. This protocol is designed to solve the energy-delay tradeoff for wireless sensor network in applications that may have critical data. Our key contribution is the usage of a deterministic scheduling for reception to insure data reception by its destination, for the packets sent in the CFP period, in the same superframe.

The simulations outcomes proved the enhancements expected by FF-MAC. These results show that FF-MAC outperforms both the IEEE 802.15.4 standard and the enhanced superframe, and provides a very low delay. The duty cycle and nodes density have no impact on the delay which make our protocol suitable for applications with heterogeneous data priorities and QoS requirements.

The presented results are encouraging and open many research perspectives. As a first step we plan to test our protocol in real world environment using iLive sensors [30]. This step is very important to validate our approach taking into account the real world impairments.

Acknowledgment

This project is supported by Franco-Moroccan cooperation program in information and communication sciences and technologies (STIC) research field, the national centre of scientific and technical research (CNRST Morocco) and the EU (FEDER).

References

[1] IEEE Computer Society, « 802.15.4 IEEE Standard for Information technology ». 2006.

[2] K. El Gholami, K.-M. Hou, et N. Elkamoun, « Enhanced Superframe Structure of the IEEE802.15.4 Standard for Real-time Data Transmission in Star Network », International Journal of Computer Applications, vol. 51, n° 15, p. 26‑32, août 2012.

[3] W. Ye, J. Heidemann, et D. Estrin, « An energy-efficient MAC protocol for wireless sensor networks », in IEEE INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings, 2002, vol. 3, p. 1567 ‑ 1576 vol.3.

[4] T. van Dam et K. Langendoen, « An adaptive energy-efficient MAC protocol for wireless sensor networks », in Proceedings of the 1st international conference on Embedded networked sensor systems, New York, NY, USA, 2003, p. 171–180.

[5] S. Mehta et K. S. Kwak, « H-MAC: A Hybrid MAC Protocol for Wireless Sensor Networks », arXiv:1003.3307, march 2010.

[6] M. Buettner, G. V. Yee, E. Anderson, et R. Han, « X-MAC: a short preamble MAC protocol for duty-cycled wireless sensor networks », in Proceedings of the 4th international conference on Embedded networked sensor systems, New York, NY, USA, 2006, p. 307–320.

[7] A. El-Hoiydi et J.-D. Decotignie, « WiseMAC: An ultra low power MAC protocol for multi-hop wireless sensor networks », in Lecture notes in computer science, p. 18‑31.

[8] S.-H. Yang, H.-W. Tseng, E. H.-K. Wu, et G.-H. Chen, « Utilization based duty cycle tuning MAC protocol for wireless sensor networks », in IEEE Global Telecommunications Conference, 2005. GLOBECOM ’05, 2005, vol. 6, p. 5 pp. ‑ 3262.

[9] J. Li, D. Zhang, L. Guo, S. Ji, et Y. Li, « M-cube: A Duty Cycle Based Multi-channel MAC Protocol with Multiple Channel Reservation for WSNs », in 2010 IEEE 16th International Conference on Parallel and Distributed Systems (ICPADS), 2010, p. 107 ‑ 114.

[10] S. Du, A. K. Saha, et D. B. Johnson, « RMAC: A Routing-Enhanced Duty-Cycle MAC Protocol for Wireless Sensor Networks », in IEEE INFOCOM 2007. 26th IEEE International Conference on Computer Communications, 2007, p. 1478 ‑ 1486.

[11] I. Rhee, A. Warrier, M. Aia, J. Min, et M. L. Sichitiu, « Z-MAC: A Hybrid MAC for Wireless Sensor Networks », IEEE/ACM Transactions on Networking, vol. 16, n° 3, p. 511 ‑ 524, June 2008.

[12] « International Society of Automation News Release, ISA Standards & Practices Board ratifies ISA-100.11a document ». .

[13] Y. Li, C. S. Chen, Y. Song, et Z. Wang, « Real-time QoS support in wireless sensor networks: a survey », in In: 7th IFAC International Conference on Fieldbuses & Networks in Industrial & Embedded Systems - FeT’2007, 2007.

[14] Y.-G. Hong, H.-J. Kim, H.-D. Park, et D.-H. Kim, « Adaptive GTS allocation scheme to support QoS and multiple devices in 802.15.4 », in Proceedings of the 11th international conference on Advanced Communication Technology - Volume 3, Piscataway, NJ, USA, 2009, p. 1697–1702.

[15] L. Cheng, X. Zhang, et A. G. Bourgeois, « GTS allocation scheme revisited », Electronics Letters, vol. 43, n° 18, p. 1005 ‑ 1006, 2007.

[16] A. Koubâa, M. Alves, E. Tovar, et A. Cunha, « An implicit GTS allocation mechanism in IEEE 802.15.4 for time-sensitive wireless sensor networks: theory and practice », Real-Time Syst., vol. 39, n° 1‑3, p. 169–204, August 2008.

[17] Y. Zhou, Y. Wang, J. Ma, J. Jia, et F. Wang, « A Low-Latency GTS Strategy in IEEE802.15.4 for Industrial Applications », in Second

International Conference on Future Generation Communication and Networking, 2008. FGCN '08, 2008, vol. 1, p. 411 - 414.

[18] A. van den Bossche, T. Val, et E. Campo, « Modelisation and validation of a full deterministic medium access method for IEEE 802.15.4 WPAN », Ad Hoc Networks, vol. 7, n^{o} 7, p. 1285- 1301, sept. 2009.

[19] B. Nefzi et Y.-Q. Song, « QoS for wireless sensor networks: Enabling service differentiation at the MAC sub-layer using CoSenS », Ad Hoc Networks, vol. 10, n^{o} 4, p. 680- 695, June 2012.

[20] H. Kim et S.-G. Min, « Priority-based QoS MAC protocol for wireless sensor networks », in IEEE International Symposium on Parallel Distributed Processing, 2009. IPDPS 2009, 2009, p. 1- 8.

[21] T. H. Kim et S. Choi, « Priority-based delay mitigation for event-monitoring IEEE 802.15.4 LR-WPANs », IEEE Communications Letters, vol. 10, n^{o} 3, p. 213 - 215, March 2006.

[22] A. Koubaa, M. Alves, B. Nefzi, et Y.-Q. Song, « Improving the IEEE 802.15.4 Slotted CSMA/CA MAC for Time-Critical Events in Wireless Sensor Networks », presented at the Proceedings of the Workshop of Real-Time Networks (RTN 2006), Satellite Workshop to ECRTS 2006, 2006.

[23] M. Youn, Y.-Y. Oh, J. Lee, et Y. Kim, « IEEE 802.15.4 Based QoS Support Slotted CSMA/CA MAC for Wireless Sensor Networks », in International Conference on Sensor Technologies and Applications, 2007. SensorComm 2007, 2007, p. 113 - 117.

[24] I. Demirkol et C. Ersoy, « Energy and delay optimized contention for wireless sensor networks », Computer Networks, vol. 53, n^{o} 12, p. 2106- 2119, August 2009.

[25] M. A. Yigitel, O. Durmaz Incel, et C. Ersoy, « Diff-MAC: a QoS-aware MAC protocol with differentiated services and hybrid prioritization for wireless multimedia sensor networks », in Proceedings of the 6th ACM workshop on QoS and security for wireless and mobile networks, New York, NY, USA, 2010, p. 62–69.

[26] E.-J. Kim, M. Kim, S.-K. Youm, S. Choi, et C.-H. Kang, « Priority-based service differentiation scheme for IEEE 802.15.4 sensor networks », AEU - International Journal of Electronics and Communications, vol. 61, n^{o} 2, p. 69- 81, February 2007.

[27] M. Al-Mamun, G. C. Karmakar, et J. Kamruzzaman, « QoS-Centric Collision Window Shaping for CSMA-CA MAC Protocol », in 2010 IEEE Global Telecommunications Conference (GLOBECOM 2010), 2010, p. 1 - 6.

[28] « The Network Simulator - ns-2 ». [Online]. Available: http://www.isi.edu/nsnam/ns/. [Accessed: 23-janv-2013].

[29] J. Zheng, Myung J. Lee, « A comprehensive performance study of IEEE 802.15.4 », Sensor Network Operations, IEEE Press, Wiley Interscience, vol. 4, p. 218- 237, 2006.

[30] H.-L. Shi, K. M. Hou, H.-Y. Zhou, et X. Liu, « Energy Efficient and Fault Tolerant Multicore Wireless Sensor Network: E^{2}MWSN », in 2011 7th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM), 2011, p. 1 - 4.

AUTHORS PROFILE

EL GHOLAMI Khalid received the B.S. degree in Physics and M.S. degree in Networking and Telecommunication from the University of Chouaib Doukkali, Morocco in 2006 and 2009, respectively. His research interest is in the area of modeling, analysis and optimization of protocols for wireless sensor networks. He is now a double degree Ph.D. student in the Department of Physics, Chouaib doukkali university, Morocco and the Department of Computer Science, Blaise Pascal university, France.

Najib EL KAMOUN received his PHD degree in Optical and Microwave Communication from the National Polytechnic Institute of Grenoble, France, in 1990. He is currently Professor Researcher at Faculty of Science, University Choaib Doukkali ElJadida Morocco. With over 20 years of expertise in information technology and communication, he has conducted several thesis and overseas missions in e-learning and telecommunication networks.His research interests include High Speed Network Architectures (MPLS), Mobility Management, security and QoS in Emerging Networks (MANET, VANET and WSN), Wireless Communications and Traffic Engineering for Computer and Telecommunication Networks.

Kun Mean HOU was born in Cambodia in 1956. He held a PhD degree in 1984 and a HDR degree in 1996 in Computer Science from the University of Technology of Compiègne (UTC). He worked as associate professor at UTC from 1984 to 1986. In 1986 he joined IN2 as R&D engineer group leader. From 1989 to 1996, he created a research group which investigated parallel architecture dedicated to real-time image processing at laboratory HEUDIASYC UMR CNRS (UTC). In 1997 he joined the college of engineering school 'ISIMA: Institut Supérieur d'Informatique de Modélisation et de leurs Applications' as professor, where he created the SMIR 'Systèmes Multisensoriels Intelligents integrés et Répartis' team of the laboratory LIMOS UMR 6158 CNRS (10 researchers) working on the development of basic hardware and software dedicated to WSN. Different sensor nodes (Bluetooth, WiFi and ZigBee), embedded wireless communication and embedded real-time kernel (SDREAM and LIMOS) are implemented and deployed in different applications such as telemedicine, intelligent transportation system and precision agriculture. He holds 3 EU patents, and he evolved in 3 EU projects and 10 technology transfers. He also evolved in several scientific committees and boards.

VHDL Design and FPGA Implementation of a Parallel Reed-Solomon (15, K, D) Encoder/Decoder

Mustapha ELHAROUSSI, Asmaa HAMYANI, Mostafa BELKASMI
ENSIAS RABAT
MAROC

***Abstract*—In this article, we propose a Reed Solomon error correcting encoder/decoder with the complete description of a concrete implementation starting from a VHDL description of this decoder. The design on FPGA of the (15, k, d) Reed Solomon decoder is studied and simulated in order to implement an encoder/decoder function.*The* proposed architecture of the decoder can achieve a high data rate, in our case, 5 clock cycles, and having a reasonable complexity (1010 CLBs).**

Keywords: Error detecting correcting codes; Reed-Solomon encoder/decoder; VHDL language; FPGA

I. Introduction

Nowadays, we live in a world where communications play an important role both in our daily lives and in their involvement in the economic and technological fields. We constantly need to increase the flow of transmission while maintaining and improving their quality. But without a concern of reliability, all improvement efforts would be futile because it would necessarily mean that some data are to be rebroadcast An error correcting code allows the correcting of one or several errors in a code word by adding redundant symbols to the information, otherwise called, control symbols.

Different possible codes exist but in this document we will only deal with Reed Solomon codes because for the moment being, they represent the best compromise between effectiveness (symbols of parity added to the information) and complexity (coding difficulty). The theory will present two decoding methods concerning Reed – Solomon codes. The first solution is the method of the Euclidean division adopted in this article, while the second method will highlight the Berlekamp-Massey algorithm.

In this work, we will present the hardware achievement of Reed Solomon encoder/decoder circuits for a (15, k, d) Code. The hardware implementation will be carried out by using programmable logic circuits of the type FPGA Altera, all translated into VHDL language. The VHDL implementation will be treated and simulated using Quartus II.

After recalling the principle of encoding/decoding of Reed-Solomon code, this paper presents the design and VHDL implementation on FPGA of (15, k, d) Reed-Solomon decoders following a pipeline and / or parallelized structure.

II. Reed Solomon Codes

A. (15, k, d) Reed-Solomon codes

The codes of Reed Solomon are non binary BCH codes belonging to the Galois fields GF ($q=2^4$). Each symbol q-areas of the body can be represented by m binary elements. The main Reed Solomon code parameters are (n, k, d) with n representing the length words of the codes, k representing the length of the information messages and d its Hamming distance. The (15, k, d) Reed Solomon code is wholly defined by the generator polynomial g (x). The primitive and irreducible polynomial is of the form $P(x) = x^4 + x + 1$. The Galois field contains 16 elements and α is a root of P(x). The generator polynomial g (x) characterizes the properties of the code. The size of the symbols is 4 bits [1], [2].

B. RS encoder

The minimal distance d allows determining the ability of correction of the error correcting codes. The parameters are defined:

- the length of the code : $n = 2^m - 1$
- the size of the message : $k = 2^m - 1 - 2*t$
  t: represents the error correction capability of the code
- The Hamming distance : $d = 2*t + 1$

The polynomial generator g (x) is defined as:

$$g(x) = \prod_{i=0}^{d-2} x - \alpha^{i}) = (x-\alpha^{0})(x-\alpha^{1})...(x-\alpha^{d-2}) \quad (1)$$

For the (15, k, d) Reed Solomon code, the information M (x) can be put in the following polynomial form:

$$M(x) = \sum_{i=k-1}^{0} \alpha_i x^i = \alpha_{k-1}x^{k-1} + ... + \alpha_1 x^1 + \alpha_0 x^0 \quad (2)$$

where $\alpha_i \in GF(16)$

The redundancy is the remainder of the division: x^{n-k} *M(x) by the generator polynomial g(x). All the additions are made in modulo two arithmetic. The rest can be written in the following form:

$$R(x) = \sum_{j=n-k-1}^{0} r_j x^j = r_{n-k-1}x^{n-k-1} + ... + r_1 x + r_0 \quad (3)$$

where $r_j \in GF(16)$

The remainder R (x) thus obtained completes the message to make the codeword C (x), in this way the literal expression of C(x) is given by:

$$C(x) = x^{n-k} \sum_{i=k-1}^{0} \alpha_i x^i + \sum_{j=n-k-1}^{0} r_j x^j \qquad (4)$$

The coding is systematic. The polynomials coefficients M (x), R (x) and C (x) can be represented either in the form of discrete values included between 0 and 15, or in the form of the power of α..

C. RS decoder

The code word C (x) transmitted may be subject to alterations due to the environment. The received code word C'(x) is equal to:

$$C'(x) = [C(x) + E(x)]\,\mathrm{Mod}\,2\,. \qquad (5)$$

E (x) represents the expression of the errors polynomial.

$$E(x) = \sum_{j=n-k-1}^{0} b_j x^j = b_{n-k-1} x^{n-k-1} + \ldots + b_1 x^1 + b_0 x^0 \quad (6)$$

Were $b_j \in GF(16)$.

III. VHDL CIRCUIT DESIGN

The FPGA have known a great improvement in size and speed. Also, the FPGA constitute a more appropriate platform for the implementation of the applications of the error detecting correcting codes.

Several studies on Reed Solomon "encoders/decoders" have already been carried out both at the university [5][6] or industrial [7][8] levels. The VHDL description of the Reed Solomon code is made so that every block of the proposed architecture is described in an independent entity.

A. Encoding

In this study, we have chosen the RS codes with parameters: (n, k, d) = (15, 9, 7) and (15, 11, 5). The circuit which instils the key encoding equation is given in figure1:

$$C(x) = M(x) * x^6 + [M(x) * x^6] \bmod g(x)$$

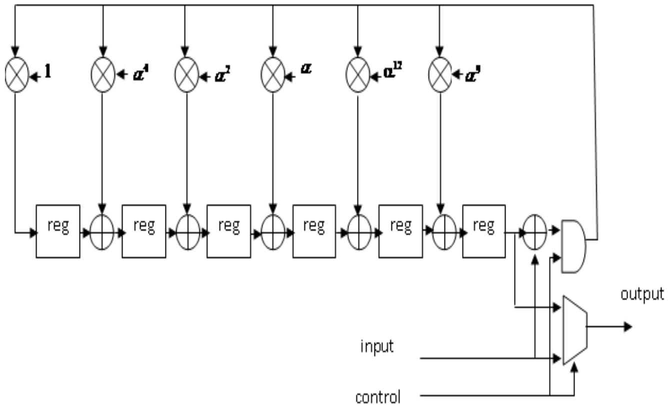


Figure 1: Encoder circuit for the RS (15, 9, 7)

The generator polynomial coefficient is given by Tab 1. The figure 2 shows the input / output pins of the RS encoder.

TAB 1: GENERATOR POLYNOMIAL COEFFICIENTS OF THE TWO RS CODES

RS Code	*Generator g(x)in linear and exponential representations*
RS (15, 9)	$g(x) = x^6 + \alpha^9 x^5 + \alpha^{12} x^4 + \alpha x^3 + \alpha^2 x^2 + \alpha^4 x + 1$ $g(x) = x^6 + 10\,x^5 + 15\,x^4 + 2x^3 + 4\,x^2 + 3\,x + 1$
RS (15, 11)	$g(x) = x^4 + \alpha^{12} x^3 + \alpha^4 x^2 + x\; + \alpha^6$ $g(x) = x^4 + 15\,x^3 + 3\,x^2 + x + 12$

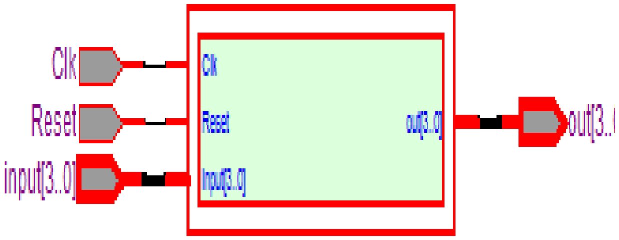


Figure 2: The input / output pins of the encoder.

Clk: Timing clock signal.

Reset: Signal allowing the reset the encoder.

Input: Input signal of symbols to encode.

Out: Output signal of encoded symbols.

Figure 3 shows the functional simulation of our (15, 9) encoder circuit, the latter has a complexity of 38 LEs.

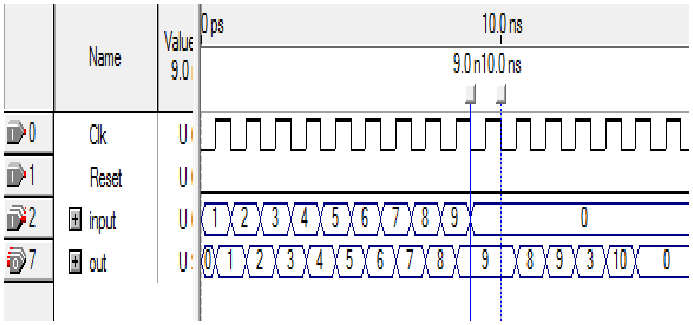


Figure 3: Simulation of the encoder (coded message)

B. Decoding

For decoding we have used the following architecture:

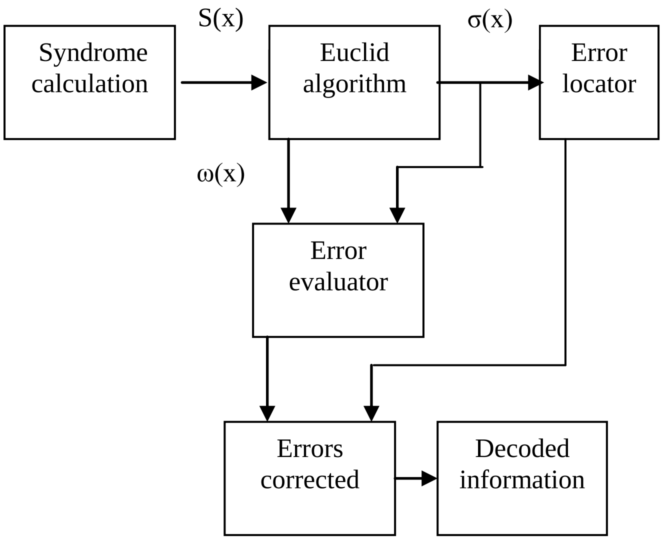


Figure 4: The Architecture Used For The (15, K, D) Reed Solomon Decoder.

- Syndrome S(x)

A syndrome for a Reed Solomon code is a polynomial with 2 * t coefficients (table 2) that depend only errors and are calculated by substituting 2 * t roots of the polynomial generator in C'(x).

$$S(x)=\sum_{i=1}^{d-1} S_i x^{i-1} = S_1 + S_2 x + S_3 x^2 + \ldots + S_{d-1} x^{d-2} \qquad (7)$$

$$S_i = C'(\alpha^{i\text{-}1}) \qquad \text{et} \quad i \in \{1, d-1\}$$

TAB 2: SYNDROME POLYNOMIAL COEFFICIENTS FOR THE RS CODES

Code RS	*S(x)*
RS(15, 9) case	$S(x)=\alpha^4x^5+\alpha^{14}x^4+\alpha^6x^3+\alpha^2x^2+\alpha^4x+\alpha^{12}$
RS(15, 11) case	$\alpha^6x^3+\alpha^2x^2+\alpha^4x+\alpha^{12}$

- Polynomial locaters and evaluators.

The decoding method of Reed Solomon codes is based on the solving of key equation:

$$\omega(x) = S(x)\sigma(x) \bmod (x^{2t})$$

where S (x): Syndrome polynomial.
σ(x) : Error locator polynomial.
$\omega(x)$ Error evaluator polynomial

The Euclid algorithm allows to calculate these polynomials. The positions of the errors are located at the roots of the error locator polynomial which are calculated by brute force using the Chien-search. The error values are then calculated using Forney algorithm**.**

The circuit represented in figure 5 reflects Euclid algorithm.

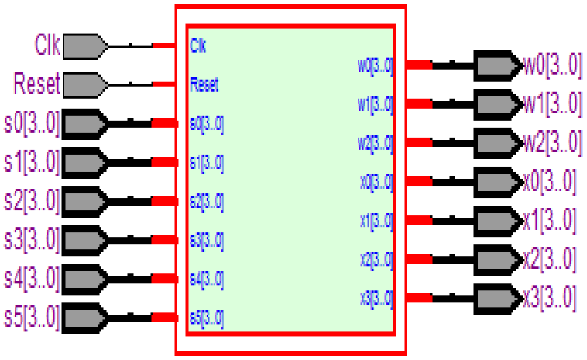


Figure 5: Euclid Algorithm Embedding Scheme.
(w: Evaluator polynomial coefficient)
(x: Locator polynomial coefficient)

- Performance of Reed-Solomon code

The Figure 6 shows the performances of RS decoders (15.9) and RS (15.11), altered by AWGN channel noise with a BPSK modulation.

IV. TESTS AND RESULTS

The methodology of simulation adopted in this work is to form the codeword C (x) using the encoding algorithm. The errors Injection consists of adding modulo 2 the codeword C(x) and the error polynomial E(x). The simulation example treats the case of a message with two errors. The codeword to be transmitted is the following sequence depending on the chosen code:

$$1, \alpha, \alpha^4, \alpha^2, \alpha^8, \alpha^5, \alpha^{10}, \alpha^3, \alpha^{14}.$$

Or $\quad 1, \alpha, \alpha^4, \alpha^2, \alpha^8, \alpha^5, \alpha^{10}, \alpha^3, \alpha^{14}, \alpha^9, \alpha^7$

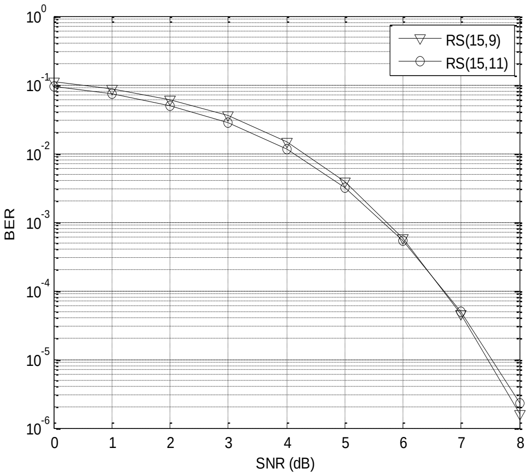


Figure 6: Performances Of The RS Decoder

The message at the output of the two encoders is shown by table 3.

Tab 3:The codeword at the output of the two encoders

Code RS	The codeword at the output of the encoders
RS(15, 9)	$1,\alpha,\alpha^4,\alpha^2,\alpha^8,\alpha^5,\alpha^{10},\alpha^3,\alpha^{14},\alpha^{14},\alpha^3,\alpha^{14},\alpha^4,\alpha^9,0.$
RS(15, 11)	$1,\alpha,\alpha^4,\alpha^2,\alpha^8,\alpha^5,\alpha^{10},\alpha^3,\alpha^{14},\alpha^9,\alpha^7,\alpha^4,\alpha^0,\alpha^6,\alpha^6.$

The received message is affected by two errors in positions 2 and 9 with the amplitudes α and α^{13} respectively.

The received message is:

$1,\alpha,\alpha^4,\alpha^2,\alpha^8,\alpha^7,\alpha^{10},\alpha^3,\alpha^{14},\alpha^{14},\alpha^3,\alpha^{14},1,\alpha^9,0.$

or $1,\alpha,\alpha^4,\alpha^2,\alpha^8,\alpha^7,\alpha^{10},\alpha^3,\alpha^{14},\alpha^9,\alpha^7,\alpha^4,\alpha^4,\alpha^6,\alpha^6.$

Or in decimal: 1, 2, 3, 4, 5, **11**, 7, 8, 9, 9, 8, 9, **1**, 10, 0.

Or 1, 2, 3, 4, 5, **11**, 7, 8, 9, 10, 11, 3, **3**, 12, 12

The figure 7 shows the syndrome coefficients and the figure 8 shows the coefficients of the two polynomials $\omega(x)$ and σ(x). The positions of the errors are located in the figure 9.

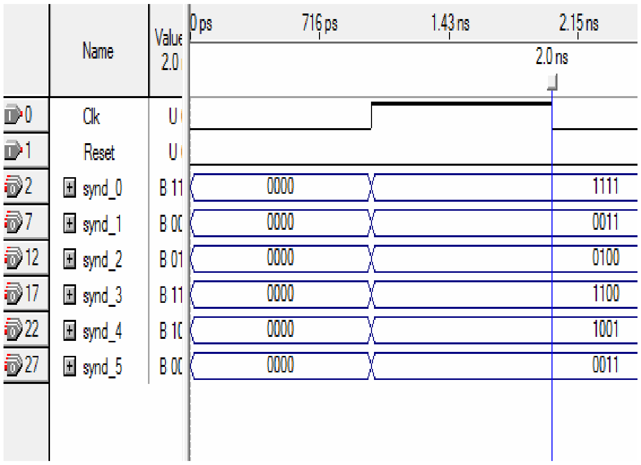


Figure 7a: syndrome of the received message of code RS(15,9)

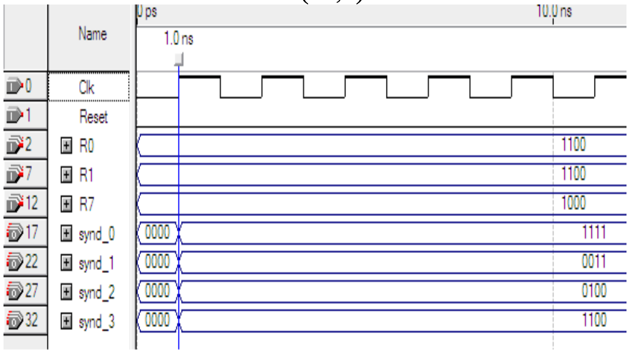


Figure 7b: Syndrome Of The Received Message Of Code RS(15,11)

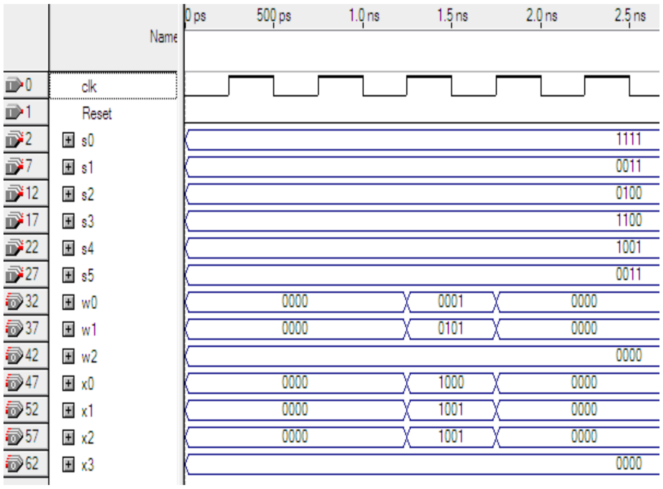


Figure 8a: Error Locator And Evaluator Polynomial
(w: Evaluator polynomial coefficient)
(x: Locator polynomial coefficient)
RS (15, 9)

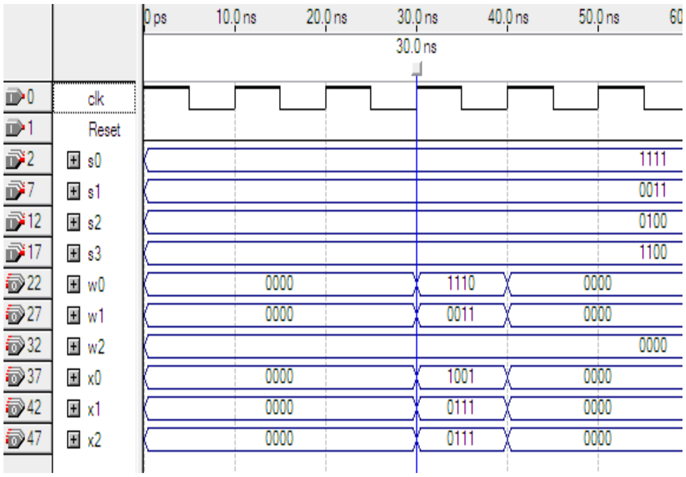


Figure 8b Error locator and evaluator polynomial
(w: Evaluator polynomial coefficient)
(x: Locator polynomial coefficient)
for RS(15,11) code

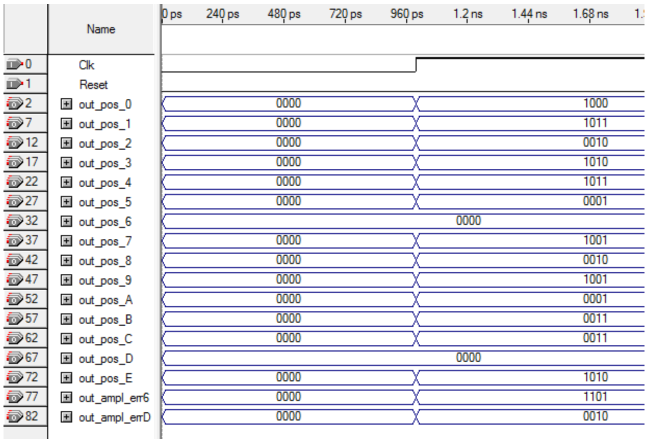


Figure 9a: Detection of the positions and the amplitudes of the two errors for RS (15, 9) code

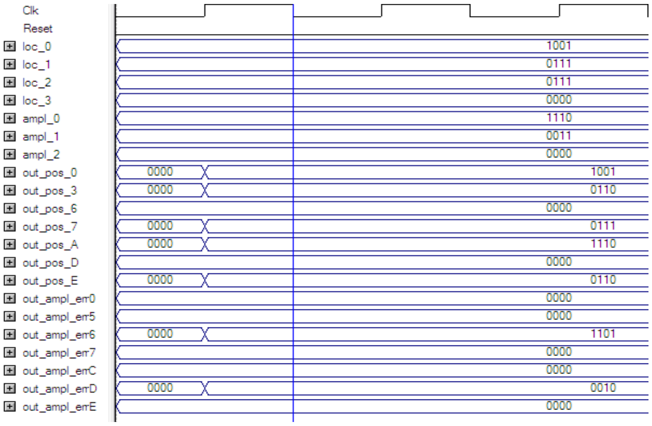


Figure 9b : Detection Of The Positions And The Amplitudes Of The Two Errors For RS(15,11) Code

The Altera's FPGA FLEX10K on which we separately tested the different blocks, from the beginning, contains 1728 LEs and 189 input / output pins. The chosen encoder/decoder architecture, presented in the previous sections, was described in VHDL and embedded on FPGA (EPF10K30RI2404) using the software Quartus II from the Altera company. The area occupied by each circuit is given in the following table 4:

TAB 4: THE AREA OCCUPIED BY DIFFERENT BLOCKS

	LEs number RS(15, 9) case	LEs number RS(15, 11) case
Syndrome computing	**175**	**114**
Euclid algorithm	**286**	**243**
Errors positions detection and Errors values calculation	**511**	**511**

The architecture chosen for this implementation reduces the number of cycles N necessary to have decoded data: $N = N_1 + N_2 + N_3$

N_1 : Number of Timing Clock cycles needed in calculating the syndrome (1 cycle in our case).

N_2 : Number of cycles to calculate σ(x) and $\omega(x)$ (3 cycles).

N_3 : Number of cycles necessary to determine the position and the correction of errors (1 cycle).

For the case of (15, k, d) Reed Solomon code the operations require a latency of 5 Timing Clock cycles. As for the architecture adopted in [1], 8 Timing Clock cycles are required.

The area occupied for the decoder for the (15, 9, 7) RS code is approximately of 972 LEs, and 868 LEs for the decoder of (15, 11) code RS. That is to say, we have reduced the area occupied in relation to the results in [1] [2] [9] [10].

V. CONCLUSION

The design of the encoder/decoder was described in VHDL and validated on FPGA (type FLEX10K30) using the software Quartus II of the company Alteras. The results showed that the area occupied and the latency is very convincing. Indeed, we have decreased the latency and the area occupied by adopting architecture in which each block is pipeline and/or parallelized.

REFERENCES

[1] S.Najah et M.Mrabti."Conception VHDL et implémentation sur FPGA du code Reed Solomon(15,k,d),traitement de signal, vol. 22, N° 2, p. 149-155.

[2] A Dandache, T Vallino, F Onteiro et J.P Delahaye "code Reed Solomon (127,k,d) avec effacement : simulation et conception sur réseaux de circuits programmables FPGA " traitement de signal, 1999, volume 16, n°4, pp 331-341.

[3] H. Lee and A. Azam " pipelined recursive modified Euclidean algorithm block for low-complexity, high-speed Reed Solomon decoder", ELECTRONICS LETTERS 18th September 2003 Vol. 39 No. 19.

[4] Hsie-Chia Chang, Ching-Che Chung, Chien-Ching Lin, and Chen-Yi Lee "A High Speed Reed Solomon decoder Chip using Inversionless Decomposed Architecture for Euclidean Algorithm", ESSCIRC, Issue, 24-26 Sept. 2002, p. 519-522.

[5] A Dabbagh "Etude et conception d'un circuit de détection d'erreurs en transmission d'informations numériques". PhD Thesis; Uuniversité de Rennes I, France (1995). (In French)

[6] S Najah "codes détecteurs d'erreurs implémentation sur des circuits de type FPGA en utilisant le langage VHDL". PhD Thesis, Faculté des sciences Dhar Mehraz, Fez, Morocco Jan. 2006 (In French)

[7] "Reed Solomon decoder". Lattice Semiconductor Corporation 2012. http://www.latticesemi.com/products/intellectualproperty/ipcores/reedsolomondecoder.cfm

[8] Aha 4011 : "10 Mbytes/sec Reed Solomon error correction device. Product specification. Advanced Hardware Architectures", 20 july 1998. http://www.datasheetarchive.com/AHA4011*-datasheet.html.

[9] A.Hikmat " Implementation of Reed Solomon Encoder/Decoder Using FPGA", Journal of Engineering and Development, Vol. 10,N°3, September 2006.

[10] B.Tiwari and M.Rajesh, "FPGA Implementation of RS Codec for digital Video Broadcasting" VSRD-IJEECE, Vol. 2, 2012, 86-77

An Analysis of Security Challenges in Cloud Computing

Ms. Disha H. Parekh,
Assistant Professor,Faculty of Computer Applications,
Marwadi Education Foundation's Group of Institutions,
Rajkot, Gujarat, India.

Dr. R. Sridaran,
Dean,
Faculty of Computer Applications,
Marwadi Education Foundation's Group of Institutions,
Rajkot, Gujarat, India

***Abstract*— Vendors offer a pool of shared resources to their users through the cloud network. Nowadays, shifting to cloud is a very optimal decision as it provides pay-as-you-go services to users. Cloud has boomed high in business and other industries for its advantages like multi-tenancy, resource pooling, storage capacity etc. In spite of its vitality, it exhibits various security flaws including loss of sensitive data, data leakage and few others related to cloning, resource pooling and so on. As far as security issues are concerned, a very wide study has been reviewed which signifies threats with service and deployment models of cloud. In order to comprehend these threats, this study is presented so as to effectively refine the crude security issues under various areas of cloud. This study also aims at revealing different security threats under the cloud models as well as network concerns to stagnate the threats within cloud, facilitating researchers, cloud providers and end users for noteworthy analysis of threats.**

Keywords— Security threats; SQL Injection; Malevolent users; Browser Security; Malicious Attacks; Data Leakage.

I. Introduction

Cloud Computing has emerged as a very well-known technique to support large and voluminous data with the help of shared pool of resources and large storage area. Zhiguo Wan et al. [36] states that "cloud computing is a new computing paradigm that is built on virtualization, distributed computing, utility computing and service-oriented architecture." Further it is added that cloud computing has emerged as one of the most significant paradigm of the IT industry and has attracted most of the industry and academia. Peter Mell and Timothy Grance, [48] have defined cloud computing as "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

Cloud computing, indeed, is a wide-ranging term that transmits hosted services over the Internet. These hosted services are generally separated into three broad categories: Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). The internet is usually represented as the "Cloud".

A cloud service is generally used by the clients as and when needed, normally on the hourly basis. This "on-demand" or "pay as you go" approach makes the cloud service flexible, where end user can have a great deal or modest of a service the way they desire at any point of time and the service is entirely administered by the provider. Noteworthy improvements in each key components included virtualization, distributed computing and also the improved access to high-speed internet facility as well as weak economy have speeded up the inflate of cloud computing rigorously.

A cloud can either be a private or a public. According to Joel Wies et al. [29], public cloud exist when a third party is offering computing resources as a service, while in a private cloud a sole user will own and operate the computing resources. Thus a public cloud sells services to any person residing on the Internet. At present, Amazon Web Services is the major public cloud provider. A private cloud is an authorized network or a data centre that provides hosted services to a restricted number of individuals. When a service provider uses public cloud resources to produce their private cloud, the result is called as a virtual private cloud. For a cloud computing, the main aim is to offer a scalable and a very easy admittance to computing resources and Information Technology services.

Cloud computing has spawned a very noteworthy interest in both academia and industry, but it is still a budding theory. In essence, it aims to combine the fiscal utility model with the evolutionary expansion of various existing advances and computing technologies. It even unites various distributed services, as well as applications and in¬formation infrastructures that consist of groups of com¬puters, storage resources and networks. Confusion exists in IT communities about how a cloud differs from existing models and how these differences af¬fect its adoption. Some visualize a cloud as a novel technical revolution, while others consider it a natural evolution of technology, economy and culture.

As cloud computing comprehends the idea of computing as an efficacy, providers are developing a mutual-shared group of configurable resources, which clients can vigorously condition and liberate according to their varying needs. Thus, both the group providers and the users would easily benefit from the reuse of computing resources and reduction in cost.

The cloud services that are implemented or those that will be implemented will always be accompanied by several threats. Knowledge about these threats shall prove to be the first step to prevent them. Hence security is the chief concern of several clients who desire to leverage cloud services. According to Kandukuri BR et al. [21] there exist some of the basic security threats that exploit the use of Cloud Computing. An easy example of this is the exercise of botnets to spread spam and malware. The other example is the application interfaces that are required to connect to cloud services especially that are developed by third parties. These interfaces must provide the user with highly protected authentication, authorization, encryption and movement monitoring mechanisms.

Cloud computing is undeniably providing with different hosted services over the internet. These hosted services are broadly classified in three different service models, namely Infrastructure as a Service, Platform as a Service and Software as a Service which have been discussed as below:

A. *Cloud Service Models:*

As known, cloud computing provides with diverse hosted services. The various service models briefly discussed earlier have further been elaborated as below, to reveal their significance with a range of security threats further in the survey:

- Infrastructure as a Service (IaaS) also referred as Resource Clouds generally provide resources which are managed and can easily be scaled up, as services to a variety of users. They essentially supply superior virtualisation capabilities. Consequently, diverse resources may be offered via a service line: Data and storage clouds have to offer a dependable access to data of a potentially large size. The success rate of data access defines the quality of these cloud servers. As infrastructure can be dynamically scaled up or down based on the need of application resources, it helps to equip multiple tenants at the same time. Moreover, the resources that are used are generally billed by the providers on the basis of the computational usage by the users.
- Platform as a Service (PaaS) supply computational resources via a platform upon which applications and services can be urbanized and hosted. In other way, it supplies all the needed resources to build an application and service via the internet, without downloading or installing it. PaaS classically makes use of fanatical APIs to organize the performance of a server hosting engine which completes and replicates the execution according to consumer requests. As each supplier exposes their own API according to the individual key potentialities, applications developed for one precise cloud provider cannot be enthused to an additional cloud host; there are though attempts to make bigger broad programming models with cloud abilities.
- Software as a Service (SaaS): It is also referred to as Application or a Service Clouds. SaaS is the model which hosts the application as a service to its various cloud users via internet. The user utilizes the software out of the box without any integration or patching up with any infrastructure. Service clouds provide an implementation of explicit business functions and business processes as per the requirement. These applications are bestowed with unambiguous cloud capabilities using a cloud infrastructure or platform rather than providing a cloud for them. Repeatedly, types of standard application software functionality are obtainable within a cloud. One of the biggest benefits of SaaS is, it helps in costing less money than actually buying the application. It provides with cheaper and reliable applications to the organization.

The three cloud services described above attract some highly significant amount of threats. This includes modification of data without proper backup, leading to data breaches or unauthorized access to sensitive data. In case of proper data backup being taken, it is vulnerable if it is not encrypted properly. Unsecured access to resources over the cloud may lead to unauthorised usage of service, platform or even an infrastructure of the provider or other users due to the associated disadvantages of virtualization as discussed in later section.

As on date only few threats have been revealed but there still exists many a more threats that are yet unsolved. As stated by Paul Hoffman and Dan Woods [30], security is one of the major chal¬lenges to the cloud and it is often a disturbing. There exist various researches dealing with security threats with either service or deployment models. While this survey majorly covers the threats that are directly influencing deployment models, service models and various network related security threats which are generally found the main cause for various security and data breaches.

The survey of related research work done on the cloud security (CS) challenges is discussed in the first half of the paper. In the second part of paper, the challenges to the CS are discussed in detail. The discussion spans the security challenges with respect to the type of deployment, service and common network issues. The next part comprises of the discussion and conclusions followed by the tables showing advantages and disadvantages of deployment models from view point of security and the vulnerabilities of different service models as appendices.

II. Related Work

Hasan Tabaki et al. [4] illustrates the unique issues of security and privacy challenges with cloud whereas Minqi Zhou et al. [24] examined cloud security and privacy issues in terms of the special relationship between the service providers and the users in a cloud. Both the works are devoid of revealing the need and importance of virtualization security.

Kresimir Popovic et al. [25], in their research work have provided a generic overview of the security issues, the requirements and the challenges that many cloud service providers' encounter. In 2011, S. Subashini et al. [26] surveyed SQL injection flaws, its cross-site scripting, non-secure storage and invalidate redirects or forwards. None of the study shows any threats of the cloud computing security with respect to deployment models of clouds. Kui Ren et al. [27] have investigated various security challenges for the public cloud without considering the threats in service models. But, most of the existing research discusses cloud security from a nonspecific viewpoint outside a cloud. None of these works talk about the threat levels in different service models (SaaS, PaaS, IaaS) from the perspective of virtualization technologies. Although, Hsin-Yi Tsai et al. [6] have examined various threats and security issues with virtualization including service models IaaS, PaaS and SaaS, they have not mentioned the impact of virtualization on cloud security with Database as a Service (DaaS). As, because virtualization is a very essential technology to cloud computing, one must account its security threats and thus develop appropriate measures and actions. Moreover, there exists much work on threats related to either service or deployment models. But, as cloud computing technology totally depends on network and internet, various threats related to network security have been accounted in this survey.

III. CHALLENGES TO CLOUD SECURITY

Security has been one of the most challenging issues for the IT executives particularly in cloud implementation. There exist numerous security anxieties that are preventing companies from captivating advantages of the cloud. Several studies, including the one by Amit Sangroya et al. [1] quote security as the primary level confront for cloud users. In this section taxonomy related to cloud computing security has been presented.

Fig. 1 represents the schematic diagram showing the hierarchy of the cloud computing, with security challenges on both the cloud computing models: Deployment and Service models and also the issues related to Networks.The classification provided above reveals various common challenges under cloud computing. The Deployment model is classified further as Private, Public and Hybrid Cloud and the security issues of the same have been exposed in common. Further, the Service model is classified into the SaaS, PaaS and IaaS briefing its security challenges in common. The security challenges with respect to network is also shown as for any internet based service, network is considered as the backbone for cloud computing.

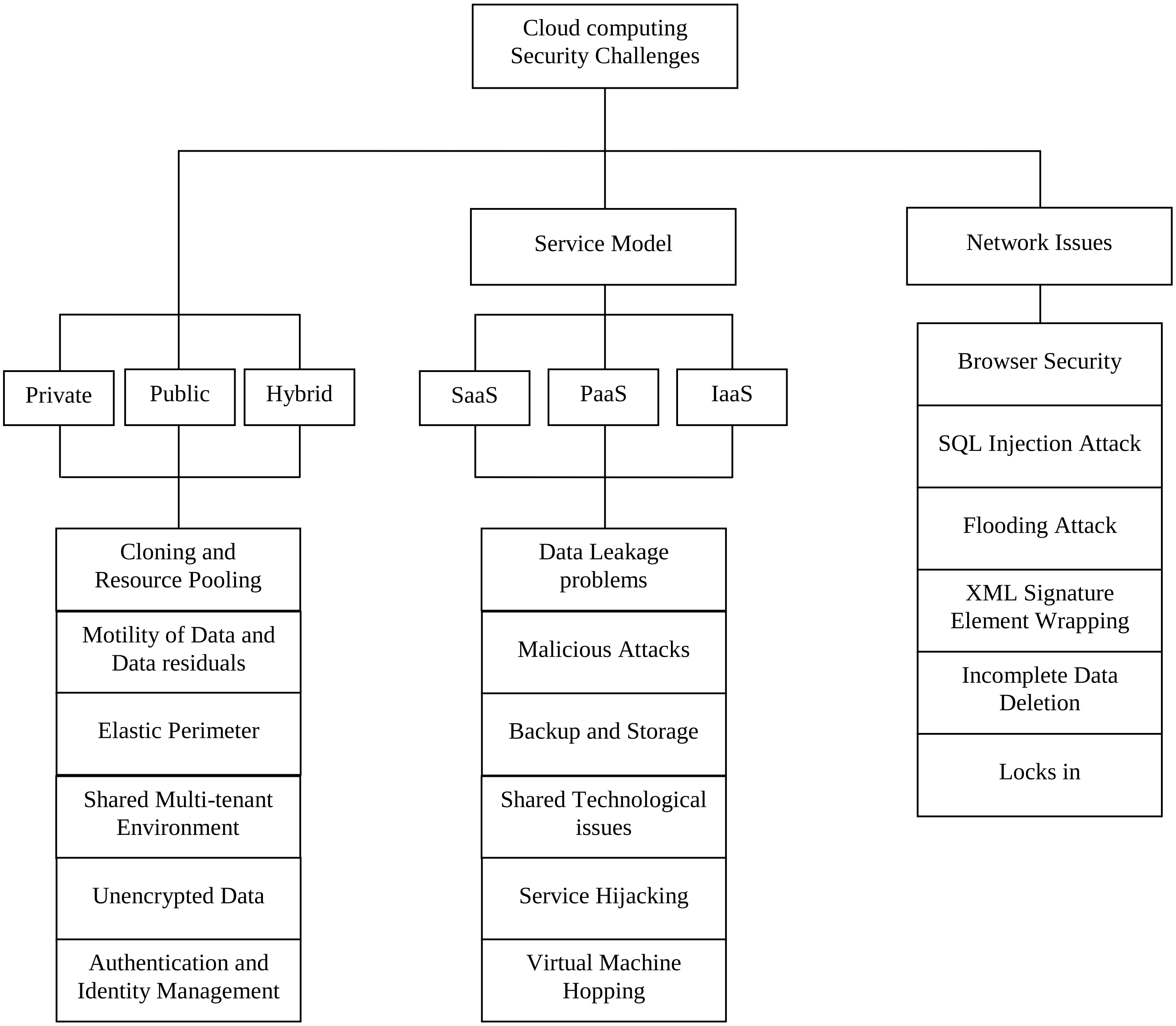


Figure 1: Classification of Security Challenge

A. Deployment Models and its security challenges:

There exist three basic types of deployment models, namely Private, Public and Hybrid clouds. Private cloud model is generally deployed within an organization and is limited only for the internal access by individuals of that organization. Public cloud model is employed by the organization for gaining access to various resources, web applications, and services over any of internet, intranet as well as extranet. Hybrid cloud is the combination of two or more clouds (public and/or private). It is an environment providing multiple service suppliers, both internal and external.

Various security challenges related to these deployment models are discussed below:

- Cloning and Resource Pooling: Cloning deals with replicating or duplicating the data. According to Bernd Grobauer et al. [2], cloning leads to data leakage problems revealing the machine's authenticity. While Wayne A. Pauley [37] describes resource pooling as a service provided to the users by the provider to use various resources and share the same according to their application demand. Resource Pooling relates to the unauthorized access due to sharing through the same network. While the study on Virtual and Cloud Computing by various researches states that a Virtual Machine can easily be provisioned, they can also be inversed to previous cases, paused, easily restarted, readily cloned and migrated between two physical servers, leading to non-auditable security threats.
- Motility of Data and Data residuals: For the best use of resources, data often is moved to cloud infrastructure. As a result the enterprise would be devoid of the location where data is put on the cloud. This is true with public cloud. With this data movement, the residuals of data is left behind which may be accessed by unauthorized users. According to Rohit Bhadauria et al. [38], data-remnant causes very less security threats in private cloud but severe security issues may evolve in public cloud donations. This again may lead to data security threats like data leakage, data remnants and inconsistent data, as stated by Hassan Takabi et al. [4]. The authors have also mentioned that in order to solve the problems with data storage the optimal solution of cryptography can be thought of effectively.
- Elastic Perimeter: A cloud infrastructure, particularly comprising of private cloud, creates an elastic perimeter. Various departments and users throughout the organization allow sharing of different resources to increase facility of access but unfortunately lead to data breach problem. In private clouds, according to Krishna Subramanian [5], the resources are centralized and distributed as per demand. The resource treatment transfers resources based on the requirements of the users thus leading to problems of data loss, where any user may try to access secure data with ease. Moreover, Marios D. Dikaiakos et al. [39] states that elasticity of various cloud based resources would lead to store replicated data on untrusted hosts and this would then lead to enormous risks to data privacy.
- Shared Multi-tenant Environment: Kui Ren et al. [27] define multitenancy as one of the very vital attribute of cloud computing, which allows multiple users to run their distinct applications concurrently on the same physical infrastructure hiding user data from each other. But the shared multi-tenant character of public cloud adds security risks such as illegal access of data by other renter using the same hardware. A multi-tenant environment might also depict some resource contention issues when any tenant consumes some unequal amount of resources. This might be either due to genuine periodic requirements or any hack attack. Hsin-Yi Tsai et al. [6], has shown that multi-tenancy makes the impact of VM Hopping attack potentially larger than conventional IT environment.
- Unencrypted Data: Data encryption is a process that helps to address various external and malicious threats. Unencrypted data is vulnerable for susceptible data, as it does not provide any security mechanism. These unencrypted data can easily be accessed by unauthorized users. According to Cong Wang et al. [40], unencrypted data risks the user data leading cloud server to escape various data information to unauthorized users. For example, the famous file sharing service Dropbox was accused for using a single encryption key for all user data the company stored. These unencrypted, insecure data, as per Marjory S. Blumenthal [32], incite the malicious users to misuse the data one or the other way.
- Authentication and Identity Management: With the help of cloud, a user is facilitated to access its private data and make it available to various services across the network. Identity management helps in authenticating the users through their credentials. But according to Rosa Sánchez et al. [35], a key issue, concerned with Identity Management (IDM), is the disadvantage of interoperability resulting from different identity tokens and identity negotiation protocols as well as the architectural pattern. While Jianyong Clien et al. [8] have mentioned that IDM leads to a problem of intrusion by unauthorized users. They even discussed that in order to serve authentication, apart from providing a password, a multi-factor authentication using smart card and fingerprint must be implemented for attaining higher level of security.

B. Service models and its security challenges:

Various cloud services like Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) are delivered and used in real time over the cloud. Bhaskar Prasad Rimal et al. [9] have mentioned SaaS as a multi tenant platform which is commonly referred to as Application Service Provider aiding distribution of services across cloud users. While the PaaS provides the developers a platform to work with all the environments and systems for the developing, testing and deploying web applications through the cloud service. The computer infrastructure needed for this application to run on a particular platform is provided by IaaS which may give more flexibility and pay-as-you-go scheme.

According to John Viegga [10], the users of SaaS have to rely heavily on the cloud provider for security purposes without any assurance to the data protection of users. In PaaS, the cloud providers offer some controls to the users building applications on their platform, without ensuring them the threats with network or intrusion prevention. While with IaaS, the developers have a better control over the application. This addresses proper security and compliance.

Various security challenges with the service models are discussed below:

- Data Leakage and consequent problems: Data deletion or alteration without backup leads to certain drastic data-related problems like security, integrity, locality, segregation and breaches. This would lead to sensitive data being accessed by the unauthorized users. As its measure provided by Rafael Moreno et al. [7], cloud platforms should provide new services in order to collect context information and to perform analysis and manage data privacy so as to support applications requesting the information. One solution to this data leakage problem, as provided by Danny Harnik et al. [13], is deduplication with allowing a limitation on number of user uploads per time window. The term deduplication means storing only a single copy of redundant data and providing just a link to this copy rather than storing actual copies of this data.
- Malicious Attacks: The threat of malicious attackers is augmented for customers of cloud services by the use of various IT services which lacks the lucidity between the procedure and process relating to service providers. Malicious users may gain access to certain confidential data and thus leading to data breaches. Farzad Sabahi [12] has shown malicious attacks by the unauthorized users on the victim's IP address and physical server. An access control mechanism tool can be thought of to control unauthorized user in accessing secured data. Peter Mell [41], has suggested Infrastructure as a Service as one of the models that exposes challenges with using virtualization as a frontier security protection to defend against malicious cloud users.
- Backup and Storage: The cloud vendor must ensure that regular backup of data is implemented that even ensure security with all measures. But this backup data is generally found in unencrypted form leading to misuse of the data by unauthorized parties. Thus data backups lead to various security threats. As per the study carried by Intel IT center [42], more the server virtualization increases, a very difficult problem with backup and storage is created. Data de-duplication is listed as one of the solution to reduce backup and offline storage volumes. But discussing about de-duplication, Danny Harnik et al. [13], have shown that de-duplication in cloud storage is carried out with the misuse of data backup.
- Shared Technological issues: IaaS vendors transport their services in a scalable way by contributing infrastructure. But this structure does not offer strong isolation properties for a multi-tenant architecture. Hence in order to address this gap, a virtualization hypervisor intercede the access between guest operating systems and the physical compute resources. As discussed by Perez R et al. [11], in spite of several advantages, these hypervisors have exhibited flaws that have permitted guest operating systems to expand inappropriate levels of control or authority on the underlying platform. This certainly led to security issues on the cloud. Lori M. Kaufman [14] has shown the implementation of IaaS by the customer to facilitate the infrastructure or hardware usage.
- Service Hijacking: Service hijacking is associated with gaining an illegal control on certain authorized services by various unauthorized users. It accounts for various techniques like phishing, exploitation of software and fraud. This is considered as one of the top most threats. According to Rajnish Choubey et al. [43], account hijacking has been pointed as one of the severe threats. The chances of hijacking ones account increases considerably as no native API's are used for registering various cloud services.
- VM Hopping: K. Owens [44] and A. Jasti et al. [45], have concluded that with VM hopping, an attacker on one VM gains rights to use another victim VM. The attacker can check the victim VM's resource procedure, alter its configurations and can even delete stored data, thus, putting it in danger the VM's confidentiality, integrity, and availability. A requirement for this attack is that the two VMs must be operating on the same host, and the attacker must recognize the victim VM's IP address. Although PaaS and IaaS users have partial authority, Thomas Ristenpart et al. [15] have shown that an attacker can get hold of or decide the IP address using benchmark customer capabilities on the basis of various tricks and combinational inputs to fetch user's IP. Thus it can be inferred that VM hopping is a rational threat in cloud computing. Additionally, multi-tenancy makes the impact of a VM hopping attack larger than in a conventional IT environment. Because quite a few VMs can run at the same time and on the same host there is a possibility of all of them becoming a victim VMs. VM hopping is thus a critical vulnerability for IaaS and PaaS infrastructures.
- VM Mobility: The contents of VM virtual disks are saved as files such that VMs can be copied from one host to another host over the system or via moveable storage devices with no physically pilfering a hard drive. VM mobility might offer quick use but could show the way to security problems likewise, the rapid spread of susceptible configurations that an attacker could make use of to endanger the security of a novel host. Several types of attacks might take advantage of weaknesses in VM mobility which includes man in-the-middle attacks. The severity of the attacks ranges from leaking perceptive information, to completely compromising the guest OS. Moreover, VM mobility augments the complication of security management because it offers augmented flexibility. In the IaaS model, a provider presents resources and underlying hardware as a service and a user can produce his or her possessed computing platform by

importing a personalized VM representation into the infrastructure service. The huge scale of IaaS makes VM mobility's force on confidentiality and integrity in the cloud possibly outsized than in a conventional IT environment. According to B. Grobauer [2], a PaaS provider offers a variety of pre-configured computing platform and solution stacks to the service users. The users take advantage of the libraries and APIs to build up their individual applications on a permanent computing platform by importing their VM images. Although PaaS considers virtualization as a key implementation technology, it does not hold up VM mobility, therefore this service model is not having the same security challenges as a traditional IT environment. While the confidentiality, integrity and availability of PaaS, SaaS and DaaS (Database-as-a-Service) are still open to the elements, the threats rose from IaaS.

- VM Denial of Service: Virtualization lets numerous VMs split physical resources like CPU, network bandwidth and memory or disk. A Denial-of-Service or DoS attack in virtualization takes place when one VM occupies all the obtainable physical resources such that the hypervisor cannot hold up more VMs and accessibility is endangered. The most excellent move towards preventing a DoS attack is to bound resource allocation using correct configurations. In cloud computing, DoS attacks could still happen as discussed by Jianyong Chen [28], but having service providers place sufficient configurations to put a ceiling on the resources owed to the VMs decreases their probability. Additionally, it is advisable to have the Service Level Agreement (SLA). This legally identifies responsibilities of the service provider and the user.

Various security threats with deployment and service models have been noticed. This makes us aware about the fact that cloud deals majorly with internet; and one need to examine various security threats with network as well. Thus certain basic issues related to network of cloud has been shown below.

C. Network issues on Cloud:

Cloud computing mainly depends upon internet and remote computers or servers in maintaining data for running various applications. The network is used to upload all the information. With the same aspect, H.B. Tabakki et al. [3] have stated security issues with network on cloud as a prime focus. It provides virtual resources, high bandwidth and software to the consumers on demand. But in reality, the network structure of this cloud faces various attacks and security issues like cloud malware injection attack, browser security issues, flooding attacks, locks-in, incomplete data deletion, data protection and XML signature element wrapping, which are explained further below.

- Browser Security: Every client uses browser to send the information on network. The browser uses SSL technology to encrypt user's identity and credentials. But hackers from the intermediary host may acquire these credentials by the use of sniffing packages installed on the intermediary host. Steve Kirsch [16] states that in order to overcome this, one should have a single identity but this credential must allow various levels of assurance which can be achieved by obtaining approvals digitally. Moreover, M. Jensen [46], has shown that Web Services security (WS-security) concept on browsers work with XML encrypted messages which does not need to be decrypted at intermediated hosts.
- SQL Injection Attack: These attacks are malicious act on the cloud computing in which a spiteful code is inserted into a model SQL code. This allows the invader to gain unauthorized access to a database and eventually to other confidential information. Further, SQL injection attacks as described by Sara Qaisar et al. [19], uses the special characters to return the data for example in SQL scripting the query usually ends up with where clause which again may be modified by adding more rows and information in it. The information entered by the hacker is misread by the website as that of the user's data and this will then allow the hacker to access the SQL server leading the invader to easily access and modify the functioning of a website. S. Roschke et al. [17] have discussed in their paper on how the network related issues hinder the cloud computing and have also shown the SQL injection attack as the top intrusion detection.
- Flooding Attacks: In this attack the invader sends the request for resources on the cloud rapidly so that the cloud gets flooded with the ample requests. As per the study carried out by IBM [18], cloud has a property to expand on the basis of large amount of request. It will expand in order to fulfil the requests of invader making the resources inaccessible for the normal users.
- XML Signature Element Wrapping: It is found to be a very renowned web service attack. According to Jamil [20], it protects identity value and host name from illegal party but cannot protect the position in the documents. The attacker simply targets the host computer by sending the SOAP messages and putting any scrambled data which the user of the host computer cannot understand. As per the studies carried out by researchers at Ruhr University, and mentioned by the editor Lee Garber [47], the XML Signature wrapping attack changes simply the content of the signed part of a message without tampering the signature. This would not let the user to understand the twisted data, thus misguiding and misleading the user.
- Incomplete Data Deletion: Incomplete data deletion is treated as hazardous one in cloud computing. According to Sara Qaisar et al. [19], when data is deleted, it does not remove the replicated data placed on a dedicated backup server. The operating system of that server will not delete data unless it is specifically commanded by network service provider. Precise data deletion is majorly impossible because copies of data are saved in replica but are not available for usage.
- Locks in: Locks in is a small tender in the manner of tools, standard data format or procedures, services edge that could embark on data, application and service portability, not leading to facilitate the customer in transferring from one cloud provider to another or transferring the services back to home IT location.

In order to effectively utilize the cloud computing technology, the research community needs to take practical and positive measures to guarantee security. According to Gunnar Petterson [22], an association exists to assume universal standards to ensure inter-operability amongst service providers. The attempts to expand security standards to warrant data's Confidentiality, Integrity and Availability, together known as CIA, are incorporated in this effort. Moreover, Kleber Vieira et al. [31] have stated that the network must be correctly trained to efficiently detect intrusions. Hence, one should count the network related issues and try to avoid them.

IV. DISCUSSION

In Table – 1 of Appendix, the security issues with the deployment models of cloud computing that is private, public and hybrid cloud are shown. Even the pros and cons of the usage of these deployed models have been listed for reference.

In Table-2 of the Appendix, the threats and the impacts of cloud service model, in terms of virtualization attacks have been compared and revealed. As discussed earlier, various virtualization vulnerabilities exist and they have a huge impact on cloud services. These virtualization vulnerabilities are discussed with respect to four types of service models basically, Software as a Service (SaaS), Database as a Service (DaaS), Infrastructure as a Service (IaaS) and Platform as a Service (PaaS). The major security challenges are bifurcated on the basis of conventional environment falling into Confidentiality, Integrity and Availability (CIA).

As indicated in III. A., cloud computing security challenges with deployment models, viz., private cloud, public and hybrid cloud have been shown along with their advantages and disadvantages at a distant, from where the security issues can be gasped at a glance. It states the advantages and disadvantages of cloud computing deployment models differentiated in three types. In accordance to their pros and cons, various security issues were concluded, showing how and which model is more secured.

From Table – 2, it is concluded that, the threats associated with VM mobility reduces with the service model PaaS. There is no direct impact of virtual mobility on SaaS, but DaaS does get affected with security challenges mainly confidentiality and integrity. Thus, although virtualization still pretences cloud computing security threats some of the description of cloud service models can hold back definite virtualization vulnerabilities.

V. CONCLUSION AND FUTURE WORK

Cloud computing has made end users both thrilled and edgy. They are excited by various opportunities provided by the cloud and are anxious as well on the questions related to the security it offers. As users migrate their data on cloud they would be alarmed with the security flaws inherent to the cloud environment. Thus security threats with cloud computing has emerged as one of the very plausible topics. This study has analyzed almost every security threat found across both the cloud models and the network and has also revealed solutions to some of them. This work will further be extended for creating a structured approach for conducting risk analysis in order to uncover security threats lying with the cloud deployed.

References

[1] Amit Sangroya, Saurabh Kumar, Jaideep Dhok, and Vasudeva Varma, "Towards Analyzing Data Security Risks in Cloud Computing Environments", Springer-Verlag Berlin Heidelberg 2010, pp. 255-265.

[2] Bernd Grobauer, Tobias Walloschek, and Elmar Stöcker, "Understanding Cloud Computing Vulnerabilities", IEEE, 1540-7993/11, 2011, pp: 50-57.

[3] H. Takabi, J.B.D. Joshi, and G.-J. Ahn, "SecureCloud: Towards a Comprehensive Security Framework for Cloud Computing Environments," Proc. 1st IEEE Int'l Workshop Emerging Applications for Cloud Computing (CloudApp 2010), IEEE CS Press, 2010, pp. 393–398.

[4] Hassan Takabi and James B.D. Joshi, University of Pittsburgh, Gail – Joon and Ahn Arizona State University, "Security and Privacy Challenges in Cloud Computing Environments", IEEE security and privacy, www.computer.org/security, 2010, pp. 24 – 31.

[5] Krishnan Subramanian, "Private, Public and Hybrid Clouds", whitepaper: Trend Micro, 2011.

[6] Hsin-Yi Tsai, Melanie Siebenhaar and André Miede, Yu-Lun Huang, Ralf Steinmetz, "Threat as a Service? Virtualization's impact on Cloud Security", IEEE, IT Pro, 2012, pp: 32- 37.

[7] Rafael Moreno-Vozmediano, Rubén S. Montero, and Ignacio M. Llorente, "Key Challenges in Cloud Computing to Enable the Future Internet of Services", IEEE, Digital Object Indentifier 10.1109/MIC.2012.69.

[8] Jianyong Chen, Yang Wang, Xiaomin Wang, "On demand Security Architecture for Cloud Computing", IEEE, 0018-9162, Digital Object Indentifier 10.1109/MC.2012.120, pp: 1 -12.

[9] Bhaskar Prasad Rimal, Eunmi Choi, Ian Lumb, "A Taxonomy and Survey of Cloud Computing Systems", 978-0-7695-3769-6/09, IEEE, pp: 44 – 51.

[10] John Viega, "Cloud Computing and the Common Man", IEEE, 0018-9162/09, pp: 106 – 108.

[11] Perez R, van Doorn L, Sailer R. "Virtualization and hardware-based security". IEEE Security and Privacy 2008;6(5):24–31.

[12] Farzad Sabahi, "Cloud Computing Security Threats and Responses", 978-1-61284-486-2, IEEE, 2011, pp: 245 – 249.

[13] Danny Harnik, Benny Pinkas, Alexandra Shulman-Peleg, "Side Channels in Cloud Services: Deduplication in Cloud Storage", 540-7993/10, IEEE, 2010, pp: 40 – 47.

[14] Lori M. Kaufman, Bruce Potter, "Monitoring Cloud Computing by Layer, Part 1", 1540-7993/11, IEEE, pp: 66 – 68.

[15] Thomas Ristenpart et al., "Hey, You, Get Off of My Cloud: Exploring Information Leakage in Third-Party Compute Clouds," Proc. 16th ACM Conf. Computer and Communications Security (CCS09), ACM Press, 2009, pp. 199–212.

[16] Steve Kirsch et al., "The Future of Authentication", 1540-7993/12, IEEE, January-February 2012, pp: 22 – 27.

[17] S. Roschke, et al., "Intrusion Detection in the Cloud," presented at the Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, Chengdu, China, 2009.

[18] Web 2.0/SaaS Security, Tokyo Research Laboratory, IBM Research. http://www.trl.ibm.com/projects/web20sec/web20sec_e.htm

[19] Sara Qaisar, Kausar Fiaz Khawaja, "Cloud Computing: Network/Security Threats and counter measures", Interdisciplinary Journal of Contemporary Research in Business, ijcrb.webs.com, January 2012, Vol 3, N0 9, pp: 1323 – 1329.

[20] Jamil, D., Zaki, H. "Security issues in cloud computing and counter measures", International Journal of Engineering Science and Technology (IJEST) , Vol. 3 No. 4, pp: 2672-2676.

[21] Kandukuri BR, Paturi VR, Rakshit A. Cloud security issues. In: IEEE international conference on services computing, 2009, p. 517–20.

[22] Gunnar Petterson, "Don't Trust and Verify: A Security Architecture Stack for the Cloud", IEEE, 1540-7993/10, pp: 83 – 86.

[23] Nir Kshetri, "Cloud Computing in Developing Economies", published by IEEE computer society, October, 2010, pp: 47 – 55.

[24] Mingi Zhou et al., "Security and Privacy in Cloud Computing: A Survey," Proc. 6th Int'l Conf. Semantics, Knowledge and Grids, IEEE Press, 2010, pp. 105–112.

[25] Kresimir Popovic and Zeljko Hocenski, “Cloud Computing Security Issues and Challenges,” Proc. 33rd Int’l Convention on Information and Comm. Technology, Electronics and Microelectronics (MIPRO 10), IEEE Press, 2010, pp. 344–349.

[26] S. Subashini and V. Kavitha, “A Survey on Security Issues in Service Delivery Models of Cloud Computing,” Journal: Network and Computer Applications, vol. 34, 2010, no. 1, pp. 1– 11.

[27] Kui Ren, Cong Wang, and Qian Wang, llinois Institute of Technology, “Security Challenges for the Public Cloud”, IEEE Press, 2012, pp. 69 – 73.

[28] Jianyong Chen, Yang Wang, Xiaomin Wang, “On demand security architecture for Cloud Computing, IEEE 2011, pp: 1 – 12.

[29] Joel Weis and Jim Alves-Foss, “Securing database as a service, issues and challenges”, Nov-Dec-2011, IEEE, pp: 49-55.

[30] Paul Hofmann and Dan Woods, “Cloud Computing: The Limits of Public Clouds for Business Applications”, IEEE, 2010, pp: 90-93.

[31] Kleber Vieira, Alexandre Schulter, Carlos Becker Westphall, and Carla Merkle Westphall, “Intrusion Detection for Grid and Cloud Computing”, Published by the IEEE computer society, IEEE, July-August 2011, pp:38-43.

[32] Marjory S. Blumenthal, “Hide and Seek in the Cloud”, IEEE, March – April 2010, pp: 57-58.

[33] Xinwen Zhang, Joshua Schiffman, Simon Gibbs, Anugeetha Kunjithapatham, Sangoh Jeong, “Securing Elastic Applications on Mobile Devices for Cloud Computing”, CCSW '09: Proceedings of the 2009 ACM workshop on Cloud computing security, pages- 127-134, November 2009.

[34] R. Buyya, C. S. Yeo, and S. Venugopa, “Marketoriented Cloud Computing: Vision, hype, and reality for delivering it services as computing utilities”, in Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications (HPCC-08, IEEE CS Press, Los Alamitos,CA, USA) 2008.

[35] Rosa Sánchez, Florina Almenares, Patricia Arias, Daniel Díaz-Sánchez and Andrés Marín, “Enhancing Privacy and Dynamic Federationn IdM for Consumer Cloud Computing”, IEEE Transactions on Consumer Electronics, Vol. 58, No. 1, February 2012, pp: 95 – 103.

[36] Zhiguo Wan, Jun’e Liu, and Robert H. Deng, “ HASBE: A Hierarchical Attribute Based Solution for Flexible and Scalable Access Control in Cloud Computing”, IEEE Transactions on Information Forensics and Security, vol. 7, no. 2, april 2012, pp: 743 – 754.

[37] Wayne A. Pauley, “Cloud Provider Transparency – An empirical evaluation”, the IEEE computer and reliability societies, IEEE, November 2010, pp: 32 – 39.

[38] Rohit Bhadauria, Rituparna Chaki, Nabendu Chaki, Sugata Sanyal, “A Survey on Security Issues in Cloud Computing and Associated Mitigation”, International Journal of Computer Applications (IJCA), June 2012, pp: 47 – 66.

[39] Marios D. Dikaiakos, George Pallis, Dimitrios Katsaros, Pankaj Mehra, Athena Vakali, “Cloud Computing – Distributed Internet Computing for IT and Scientific Research”, IEEE Internet Computing, 2009 IEEE, pp: 10 – 13.

[40] Cong Wang, Ning Cao, Kui Ren, Wenjing Lou, “Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data”, IEEE transactions on parallel and distributed systems, IEEE, Digital Object Indentifier 10.1109/TPDS.2011.282, 2011, pp: 1 – 14.

[41] Peter Mell, “What’s Special about Cloud Security?” , IEEE, IT Pro July/August 2012, pp: 6 – 8.

[42] Intel IT Center, “Preparing your Virtualized Data Center for the Cloud”, pp: 1 – 20.

[43] Rajnish Choubey, Rajshree Dubey, Joy Bhattacharjee, “A Survey on Cloud Computing Security, Challenges and Threats”, International Journal on Computer Science and Engineering, ISSN: 0975-3397, Vol. 3 No. 3 March 2011, pp: 1227 – 1231.

[44] K. Owens, “Securing Virtual Computer Infrastructure in the Cloud,” white paper, Savvis Communications Corp., 2009.

[45] A. Jasti et al., “Security in Multi-Tenancy Cloud,” Proc. IEEE Int’l Carnahan Conf. Security Technology (ICCST 10), IEEE Press, 2010, pp. 35–41.

[46] M. Jensen, “On Technical Security Issues in Cloud Computing”, IEEE International Conference on Cloud Computing, pp: 109 – 116.

[47] Lee Garber, “Serious Security Flaws identified in Cloud Systems”, News Briefs, IEEE, December, 2011, pp: 21 – 23.

[48] Peter Mell and Timothy Grance, The NIST Definition of Cloud Computing, U.S. National Institute of Standards and Technology (NIST), Special Publication 800-145, September 2011

AUTHORS PROFILE

1. Ms. Disha H. Parekh, MCA, PGDBA (Human Resource), is presently an Assistant Professor, Faculty of Computer Applications, Marwadi Education Foundation’s Group of Institutions, Rajkot, Gujarat. She has done MCA from Ganpat University, Gujarat. She has completed PGDBA. with specialization in HR from Symbiosis University. She has published 2 papers in the International Journal and has presented 1 paper at National conference. She has attended many workshops and Seminars. Her areas of interest are Software Engineering and Web Technologies. She is currently pursuing Ph.D in Bharathiyar University. She may be reached at disha.hparekh213@gmail.com

2. Dr. R. Sridaran has done his post graduation in Computer Applications and Management. He has been awarded the Ph.D in Computer Applications in 2010. Having started his career as an Entrepreneur, he has offered his consultancy services to various service sectors. He has also designed and delivered various training programs in the areas of IT & Management. He has published 14 research papers in leading Journals and Conferences and presently guiding four research scholars. He has got 17 years of academic experience and served in leading educational institutions at different capacities. He is currently the Dean, Faculty of Computer Applications, Marwadi Education Foundation’s Group of Institutions, Rajkot, Gujarat. He may be reached at sridaran.rajagopal@gmail.com

APPENDIX

TABLE I. PROS AND CONS OF DEPLOYMENT MODELS WITH THEIR LEVEL OF SECURITY

	Advantages and Disadvantages
Private Cloud	• Most control over data and platform • Latent for multi-tenancy of business units to cause fulfilment and defence risk • May not have convulsive abilities when added performance or capacity is required
Public Cloud	• Likely for better cost savings if infrastructure owned and controlled by public providers • Failure of control and data loss and platform • Possible for multi-tenancy with former organizations to reason security risk • Third party safety controls possibly not clear and may cause unidentified risks.
Hybrid Cloud	• Potential for difficulty to cause unfamiliar vulnerabilities and indefinite risks

TABLE II. VIRTUALIZATION VULNERABILITIES WITH SERVICE MODELS OF CLOUD

Cloud Computing Environment and Security Conventional Environment		**Virtualization (VM) Vulnerabilities**		
		VM Hopping	**VM Mobility**	**VM Denial-of-Service attacks**
Software as a Service	**Confidentiality**	O	✘	✘
	Integrity	O	✘	✘
	Availability	O	✘	✘
Database as a Service	**Confidentiality**	✓	✘	✘
	Integrity	✓	✘	✘
	Availability	✘	✘	✘
Infrastructure as a Service	**Confidentiality**	✓	✓	✘
	Integrity	✓	✓	✘
	Availability	✓	✓	✘
Platform as a Service	**Confidentiality**	✓	✘	✘
	Integrity	✓	✘	✘
	Availability	✓	✘	✘

O →Indicates optional. Direct impacts are not possible, but indirect impacts may cause vulnerability.

✘ → Indicates reduced occurrence of the vulnerability.

✓ → Indicates presence of vulnerability.

Applicability of Data Mining Technique using Bayesians Network in Diagnosis of Genetic Diseases

Hugo Pereira Leite Filho
University State of Goiás: Information Systems:
Ceres, Brazil

Abstract— **This study aims to identify a methodology to aid in the identification of diagnosis for chromosomal abnormalities and genetic diseases, presenting as a tutorial model the Turner Syndrome. So, it has been used classification techniques based in decision trees, probabilistic networks (Naïve Bayes, TAN e BAN) and neural MLP network (Multi-Layer Perception) and training algorithm by error retro-propagation. Described tools capable of propagating evidence and developing techniques of generating efficient inference techniques to combine expert knowledge with data defined in a database. We have come to a conclusion about the best solution to work out the show problem in this study that was the Naïve Bayes model, because this presented the greatest accuracy. The decision - ID3, TAN e BAN tree models presented solutions to the indicated problem, but those were not as much satisfactory as the Naïve Bayes. However, the neural network did not promote a satisfactory solution.**

Keywords—Turner Syndrome; probabilistic networks; classification techniques based in decision trees

I. Introduction

This study includes processing knowledge concerning data storage and manipulation by the machine so as to be used for troubleshooting, specifically, as an adjuvant in cytogenetic diagnosis of genetic diseases. Furthermore, it aims to apply the knowledge of cytogenetic and informatics to improve the quality of genetic diagnosis.

The study presents a numerical approach to treat uncertainty by the calculation of probabilities, the reasoning being based on probabilistic inferences, i.e, the calculation of the conditional probability of an event, given the available evidence and applying the Bayes Theorem.

In general there are several evidences and direct application of this numerical approach which is questionable facing the problems of complexity for big-sized applications, as it requires a huge array of conditional probabilities are estimated and provided for the system, preventing the acquisition of knowledge and implying high time requirements, storage, capacity and computing power to process information of interest [6].

It is in the outlined context above that this study fits in, proposing the use of machine learning algorithms to extract knowledge models - probabilistic networks - from databases - so this new information may help the expert in the acquisition stage of knowledge that make up the process of building a supporting system to decision making [5].

II. Probabilistic Distribution Using Bayesian Network

The Bayesian odds are often used in statistical inferences to specify a previous knowledge and combine this knowledge with the data available through Bayes Theorem [24]. Bayes's formula then provides a rule by updating previous probabilities, retrospectively when the data are known and analyzed. Given a set of data for evaluating the fitness, the best result (the fittest) is defined as the most likely model data, respecting previous knowledge of the problem understanding [3].

The probabilistic distributions occurring in uncertain causal relationships within a problem domain. The probabilistic reasoning runs on these relationships using the Bayes theorem, which can be expressed as follows [7]:

$$P(Xi \mid Y) = \frac{P(Xi)P(Y|Xi)}{P(X1)P(Y|X1) + P(X2)P(Y|X2) + P(X3)P(Y|X3) + \ldots + P(Xn)P(Y|Xn)}$$

$$\text{Para } i = 1, 2, 3, \ldots, n$$

Probabilistic inference are used in the propagation algorithms of beliefs in Bayesian networks, and can be either causal, that part from the causes for the purpose; diagnostic parting from the effects to causes; inter-cause when discriminate between the causes of a common effect or mixed, characterized by the combination of two or more types previously mentioned [15].

In probabilistic inferences, we calculate the probability of an event, given the evidence found on the network.

III. State of the art technique in AI and Data Mining

The primary objectives of the research are to develop efficient inference techniques for use in information system, for which it is necessary the availability of valid knowledge model [4]. As a result, rose up techniques of data mining and methods of knowledge based on Bayesian networks technology to extract valid knowledge models from the database.

For [4], to generate knowledge Bayesian models, many algorithms found in the literature require the following information: a) the list of variables, b) the arrangement of these variables, and c) Independence and dependency relations between them, i.e., the causal relationships. With this information passed to the learning software, it is assumed that the user has a deep knowledge of the network to be generated, an assumption that is not always true.

A. Knowledge Discovery in Database - KDD

The terminology knowledge discovery in databases was proposed in the first KDD workshop in 1989 to emphasize that the final product of the process of discovery in database was knowledge [25]. KDD has become, then, a specific interdisciplinary field that emerged in response to the need for new approaches and solutions to enable the analysis of large and complex databases [26].

The solution in data organization can be applied in the construction of Data Warehouse (DW), which allows for storing information, previously dispersed, through the identification, understanding, integrating and aggregating of the data in order to position them in the most appropriate location, to meet the organizational strategy of the company [13]. For the extraction of knowledge from the organized data are used mining tools known as data (MD), which may incorporate statistical techniques, probabilistic inferences and / or of AI, capable of providing responses to discover new knowledge in large databases.

B. Methodology CRISP-DM

The Daimler Chrysler , SPSS and NCR industries created the CRISP-DM consortium - Cross Industry Standard Process for Data Mining and proposed a reference model for the process of data mining [19] non-proprietary and available free of charge [16]. Figure 1 describes the steps of the CRISP-DM methodology that are highly representative and used in the world market.

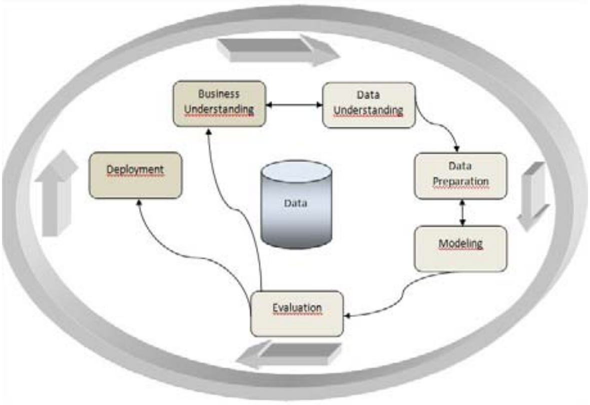


Figure 1. The process model forseen by the consortium CRISP-DM can be summarized through the lifecycle of the data mining process.

IV. Tools and Techniques Used

A. UnBMiner®

As [16], UnBMiner® is a visual and interactive environment, programmed in Java, platform independent, for carrying out the process of data mining. The framework automates much of the process CRISP-DM. More specifically, it covers part of the data preparation phase (pre-processing module UnBMiner®), Phase modeling (modeling formalisms based on Naïve Bayes, CNM, decision tree - algorithms ID3 and C4.5 - and multilayer neural network with backpropagation) and evaluation phase (evaluation module of UnBMiner®).

B. UnBBayes®

As [14], the UnbBayes is a visual and interactive environment for editing and compilation of Bayesian networks (BN)

UnBBayes is developed in Java and documented with Javadoc e JavaHelp. The system is distributed freely for non-commercial use, under license GNU GPL[1]

The software UnbBayes supports the learning of the topology and the probabilities of Bayesian networks with application of search algorithms and scoring or analysis of dependencies in large databases.

C. Classifiers based on Bayesian Networks

For [10] the network topology Naïve Bayes is represented by a tree where the root unit height is the class variable and the leaves are variable attributes.

The learning algorithm of the classifier TAN [18] first builds a structure of a tree with the variables in X \ {“TURNER”}; running test where they propagated information of mutually conditioned "TURNER". After add a link of class-node “TURNER” for each attribute-node, a structure similar to that of Naïve Bayes (i.e., the class-node is a father to all attribute-node).

The classifier BAN is an extension of the classifier TAN allowing attributes give a form the arbitrary, acyclic graph and oriented [18].

V. Data Collection

Data collection was performed and the cases were treated in LaGene/SULEIDE/SES-GO and characterized a measurement of part of a population. The statistics calculated from the samples were used to predict various population parameters [20].

Data were collected and analyzed by the technical specialist on request of laboratory tests. Although the number of intense laboratory requests, we identified two realities that hamper data collection consistency.

The first difficulty is in cases that were indicated as "clarify the case" because the attributes were not identified what determines their exclusion.

The second difficulty was the case where the technical expert identified the attributes, thus, not formed a correlation between the attributes and their classes, in these cases there was also data deletion. Therefore, were identified 84 cases where the class is TURNER.

A. Signs and Symptoms of Turner Syndrome

The Turner Syndrome (TS) occurs in approximately 1:2130 live births female [17; 11] and is due to the presence of one X chromosome and partial or total loss of the second sex chromosome. Despite the wide clinical polymorphism, it is

[1] General Public License (http://www.gnu.org/licenses/gpl.html)

considered that currently short stature, sexual infantilism and peripheral lymphedema are the striking clinical findings in ST. Its clinical sign most obvious and easily observed is short stature [2], which ranges on average between 142 and 146.8 cm [12; 8].

The gonadal dysgenesis is also an important signal in the ST, leading to secondary signs as primary amenorrhea, delayed pubertal development and infertility [1].They may also be subject to some congenital anomalies, including cardiovascular and renal problems, hearing loss, osteoporosis, obesity trend, and hypertelorism nipples.

Large variability of dysmorphic signs are also observed, as short neck and / or winged, broad chest and shield, cubitus valgus (from the Latin cubitus valgus), low implantation of the hair at the nape, prominent ears and low-set, fingernails hipoplástics, strabismus, epicanthal folds, among others [17; 9; 22; 21].

When there is clinical suspicion of ST, the appropriate test to confirm the diagnosis is the G-band karyotype, which allows the identification and analysis of chromosome. For karyotyping runs were cultured lymphocytes, in short-term (72h) obtained from the peripheral blood of the patient. Approximately 50% of cases, the notation classical karyotype is 45, X, which includes monosomy X , as shown in Figure 2.

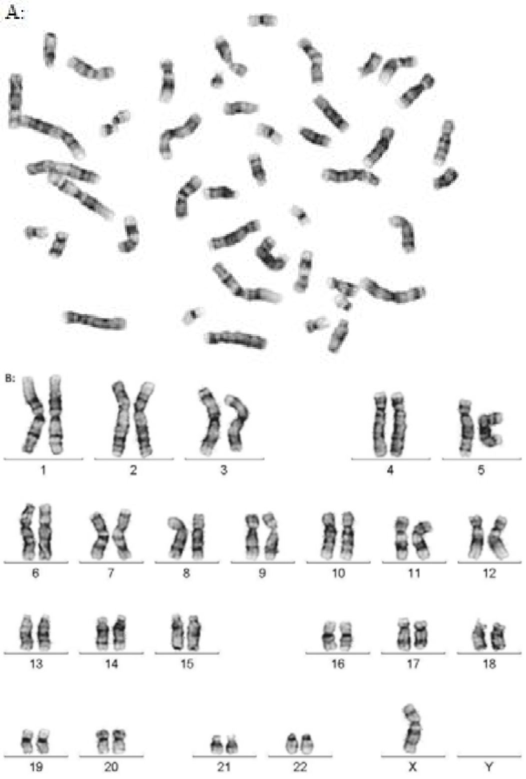


Figure 2. Human karyotype showing monosomy X, the most common chromosomal abnormality observed in Turner syndrome. A: Metaphase chromosomes containing spread, which were obtained after culturing lymphocytes for 72 hours and stained by GTG (Giemsa Trypsin +) B: chromosome pairing indicating the absence of a sex chromosome, resulting in the notation karyotype 45, X, corresponding the cytogenetic diagnosis of Turner syndrome. Source. Cytogenetics Laboratory, Human Molecular Genetics (LaGene), photograph kindly provided by Dr. Claudio C. Silva.

VI. Result and Discussion

The proposed objective of the study was to follow the steps of the methodology CRISP-DM, therefore, it is necessary attention to the attributes for the scope of our study. Based on data collected by the technical expert, the attributes were identified (signs / symptoms) and class (syndrome) of the problem.

Using UnBMiner® was possible to perform the steps defined by the CRISP-DM. The first step was made by preprocessing, where the attributes were identified.

Both the class and attributes received values "YES" or "NO" and converted to "1" and "0" respectively. We used a class: "TURNER". The class TURNER was diagnosed by 9 independent attributes, represented by: "female, short stature, chest shield, gonadal dysgenesis, hypoplastic nails, cubitus valgus, winged neck, hypertelorism nipples and a tendency to obesity." There were 84 patient records for cases of ST.

It was necessary to clean the data, thus we also used a regression using the step-wise method to perform the selection of variables and identify attributes that are relevant to the class. As the number of cases was small, so it was necessary to apply methods to validate the data with consistent results, in this case the method of choice was 4-fold cross-validation [23].

There has been chosen an algorithm and a tool – UnbBayes – able to propagate evidence and develop efficient inference techniques able to originate appropriate techniques to combine the expert knowledge with defined data in a databank.

VII. Conclusions

The present study demonstrated that the theory involving the Bayesian network has provided consistent results that allow the construction of knowledge-based systems. Thus, it was possible to integrate learning and propagation of evidence, yielding good results.

The problem encountered in building the Bayesian network and the forming of an expert probabilistic system was to obtain knowledge, in that, most data which served as a parameter for obtaining the results contained incomplete information, making the number of cases useful for the experiment were reduced.

A conclusion about the best solution to work out the shown problem in this study that was the Naïve Bayes model, because this one presented the greatest accuracy. The decision - ID3, TAN e BAN tree models presented solutions to the indicated problem, but those were not as much satisfactory as the Naïve Bayes. However, the neural network did not promote a satisfactory solution.

References

[1] AM. Pasquino, F. Passeri, I. Pucarelli, M. Segni, G. Municchi, Spontaneous pubertal development in Turner's syndrome. Italian Study Group for Turner's Syndrome. J Clin Endocrinol Metab. p. 1810–1813, 1997.

[2] BM. Lippe, Turner Syndrome. In: Sperling MA, ed. Pediatric Endocrinology. Philadephia:WB Saunders Company. p. 387-421, 1996.

[3] B-T. Zhang, A Bayesian framework for evolutionary computation. Proceedings of the 1999 Congress on Evolutionary Computation (CEC99), 1:722-728; 1999.

[4] C. Koehler, MR. Vicari, CD. Flores, MS. Nassar, Mineração de Rede bayesianas a partir de Base de Dados Médicos: Proposta de Algoritmo. Univ. de Caxias do Sul (UCS), Caxias do Sul. Brasil; 2004.

[5] C. Koehler, SM. Nassar, Modelagem de Redes bayesianas a partir de Dados Médicas. Symposio de Informática y Salud; 2002.

[6] CD. Flores, Fundamentos dos Sistemas Especialistas. Porto Alegre, Rio Grande do Sul, 2002.

[7] E. Rajabally, P. Sen, S. Whittle, J. Dalton, Aids to Bayesian Belief Network Construction Second IEEE Internattonal Conference On Intelligent Systems; June 2004. p. 7803-8278.

[8] GG. Massa, M. Vanderschueren-Lodeweyckx, Age and height at diagnosis in Turner syndrome: influence of paternal height. Pediatrics, p. 1148-52, 1991.

[9] J. Batch, Turner syndrome in childhood and adolescence. Best Pract Res Clin Endocrinol Metab. p. 465-82, 2002.

[10] J. Cheng, R. Greiner, Comparing Bayesian Network Classifiers, Proceedings of the fifteenth international conference on uncertainty in artificial intelligence, 1999.

[11] J. Nielsen, M. Wohlert, Chromosome abnormalities found among 34,910 newborn children: results from a 13-year incidence study in Arhus, Denmark. Hum Genet. p. 81–83, 1991.

[12] JG. Hall, DM. Gilchrist, Turner syndrome and its variants. Pediatr Clin North Am. P.1421-44, 1990.

[13] M. H. Brackett, The data warehouse challenge: taming data chaos. New York: John Wiley & Sons, 1996.

[14] M. Ladeira, DC. da Silva, FJF. Lima Jr., MS. Onishi, RN. Carvalho, WT. da Silva, Ferramenta Aberta e Independente de Plataforma para Redes Probabilísticas; 2002. M. Ladeira, Diagrama de Influências Múltiplo Secionado. [Tese]. Porto Alegre:[s.n], 2000.

[15] M. Ladeira, MHP. Vieira, H.A Prado, RM Noivo, DBS. Castanheira, UnBMiner® – Ferramenta Aberta para Mineração de Dados, Univ. de Brasília; 2005.

[16] MVN. Lipay, B. Bianco, ITN Verreschi, Gonadal dysgenesis and tumors: genetic and clinical features. Arq Bras Endocrinol Metab. vol. 49, nº 1. p. 60-70, 2005.

[17] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian Networks Classifier. Machine Learning, 29, p. 131-161; 1997.

[18] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth, CRISP-DM 1.0: Step-by-step data mining guide. SPSS, 1999.

[19] R. Larson, B. Farber, Estatística aplicada; trad. Cyro de Carvalho Patarra – São Paulo: Prentice Hall, 2004.

[20] [21] RG, Rosenfeld, Turner syndrome: a guide for physicians. California:The Turner Syndrome Society; 1992.

[22] RN. Rieser, LE. Underwood, Turner Syndrome: a guide for families. California:The Turner Syndrome Society; 1992.

[23] RR. Bouckaert, E. Frank. Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. Ed. Springer Berlin/Heidelberg. vol.3056/2004, p.3-12, 2004.

[24] SJ. Press, Bayesian Statistics: Principles, Models, and Applications, Wiley; 1989.

[25] UM. Fayyad, G. Piatetsky-Shapiro, P. Smyth, Knowledge Discovery and Data Mining: Towards a Unifying Framework. Second International Conference on KD & DM. Portland, Oregon; 1996.

[26] W. Romão, Descoberta de Conhecimento Relevante em Banco e Dados sobre Ciência e Tecnologia. [Tese]. Univ. Federal de Santa Catarina, Santa Catarina-Brasil; 2002.

AUTHOR PROFILE

Hugo Pereira Leite Filho received the master degree in Environmental Health Sciences, with emphasis in AI and Bayesians Networks in 2006. He is actually teaching in the institute of Information System in University State of Goiás – Brazil.

Green ICT Readiness Model for Developing Economies: Case of Kenya

Mr. Franklin Wabwoba,
Department of Information Technology
Kibabii University College

Dr. Stanley Omuterema,
Department of Disaster Management and Sustainable Development
Masinde Muliro University of Science and Technology

Dr. Gregory W. Wanyembi,
Department of Computer Science
Masinde Muliro University of Science and Technology

Mr. Kelvin Kebati Omieno
Department of Computer Science
Masinde Muliro University of Science and Technology

***Abstract*— There has been growing concerns about the rising costs of doing business and environmental degradation world over. Green ICT has been proposed to provide solutions to the two issues yet it is not being implemented fully in developing economies like Kenya. For its implementation, it is critical to establish the level of green ICT readiness of organisations to inform where to start and where to put more emphasis. Over the past few years this has been done using Molla's G-readiness model. However this model assumes the basic level of G-readiness to be same for both developed and developing economies to be the same with regard to ICT personnel preparedness. Based on green ICT readiness in Kenya, the relationship between ICT personnel's gender, age and training with the G-readiness variables as proposed in Molla's G-readiness model was investigated. The study surveyed ICT personnel in four cases using a questionnaire on a seven scale likert scale. It established that there exists a significant relationship between the ICT personnel related variables and the G-readiness variables. Based on the findings on the relationship, the study extended Molla's G-readiness model to include a sixth dimension of personnel readiness.**

Keywords— Developing economies; Extended G-readiness model; Green ICT; G-readiness model; Green ICT readiness

I. Introduction

This Organisation preparedness to implement green ICT may require answering a number of questions to determine the likelihood of doing it successfully. As noted by Molla & Cooper (2009) [1], in an environment where other change initiatives are perceived to have been well-managed, people were as involved as they wanted to be, communication infrastructure and process are already in place, with previous change projects being seen as successful and with clear success factors having been articulated, green ICT is likely to be implemented successfully.

Green ICT readiness is considered to be an organization's capability to embed sustainability in the beliefs and attitudes in the development, deployment and disposal of ICT technical assets and in their ICT processes, practices and policies and in the governance systems to ensure compliance with internal and external sustainability expectations [2]; [3]. Therefore green ICT readiness or G-readiness is demonstrated through the combination of attitude, policy, practice, technology and governance in applying environmental criteria to its ICT technical infrastructure across the key areas of ICT sourcing, operations and disposal to solve both ICT and non-ICT by using IT related sustainability problems [4]. This however cannot be achieved without the ICT personnel's valuable contribution.

II. An Overview Of Molla's G-Readiness Model

The G-Readiness model's by Molla, Cooper, & Pittayachawan (2009) [4] was adopted for this study. Green ICT readiness in the study therefore was addressed from five dimensions namely: - attitude, policy, practice, governance and technology but following analysis of personnel factors an extension of a sixth dimension of the human capability was introduced. The initial model approach is given in the Figure 1.

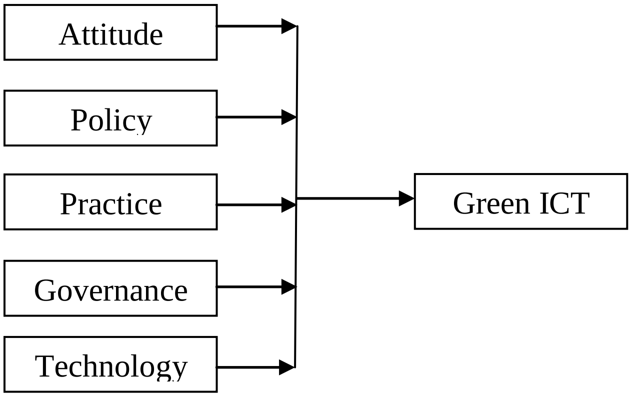


Figure 1: Green ICT readiness model
Source: Molla, Cooper, & Pittayachawan (2009) and Cooper & Molla (2012)

A. Green ICT Attitude

Attitude is an intangible thing. It describes how we think, rather than how we act, and it is about attitude or culture [5]. Green ICT Attitude deals with the extent to which ICT people and management are aware of and concerned about climate change and the environmental impact of ICT artifacts and operations [6]. The Green ICT Attitude does reflect the

We appreciate the National Council of Science and Technology (Kenya) for funding the PhD research "Enhancing green ICT through ICT personnel alignment in developing economies" from which this paper was developed.

disposition of the ICT people and management within any given organisation to the issue of climate change and to the anticipated role of ICT in organizational strategy to respond to such change towards sustainability. It may define the level of ICT professionalism [7]. Naturally this is reflected in the ICT organizations effort to improve the energy efficiency in managing the ICT technical infrastructure with a view of reducing greenhouse gas emissions and e-waste.

Attitude in this study was considered to be an organisation's ICT people sentiments, values and norms towards climate change and eco-sustainability and ICT's role [2]. Having a positive attitude towards Green ICT is very important – it precedes everything else [5]. And, as is often the case in business, those attitudes are most effective if they come from the managerial human infrastructure. "Managerial human personnel support is an essential part of any Green ICT program [1]. Attitude has been established to be one of the major factors that influence the acceptance and use of technologies hence will most likely have a major influence on the implementation of green ICT. Please do not revise any of the current designations.

B. Green ICT Policy

Green ICT Policy encompasses the frameworks an organization has developed and put in place to apply environmental sustainability criteria throughout its value chain including ICT sourcing, ICT operations and services and ICT end-of-life management [6]; [8]. Policy defines the administration of green ICT initiatives, the allocation of budget, and other resources as well as the metrics for assessing its impact [2]. A policy development framework includes the establishment of policies, the communication of those policies, the enforcement of those policies, and the measurement of policy effectiveness and mitigation strategies [5]. A green ICT policy takes into account the required roles and responsibilities, skill-sets, commitments, targets, deliverables and methodologies used. The policies give indication on the organisation's commitment to technology redundancy and the roll-over in gaining benefits of each technology advance that is taking place rapidly.

It does encompasses the frameworks an organisation puts in place to apply environmental criteria green ICT activities and operations [4]. This measure determines how organisations have advanced their green ICT initiatives. Green ICT readiness policies are those that discourage environmental negative impacts such as energy intensive production methods, wasteful packaging, poor recycling practice and heavy use of hazardous materials use and practices. The main areas for assessment of policy readiness according to Molla and Cooper (2009) are IT sourcing, IT operations and services and IT end of life management.

Some operations and services policy considerations are inclusive of computer power management, computer use policies for the staff. This includes of policies that guide the extent to which services provided by ICT support issues of sustainability [1].

IT end of life policy are policies and regulations that guide disposal and settlement of ICT infrastructure. They have guidelines on how to deal with end life of equipment such as how to recycle them or how to do away of the same.

It is essential for organisations to have and follow the National Plan and National ICT policy as well as organisational policy on green ICT. In so doing they will have set of their goals to employ green ICT in the future. Such a plan may positively affect enhancement of green ICT because the human personnel especially the ICT trainers may prepare themselves and learners for the future by changing their behaviour so as to increase the utilization green ICT.

C. Green ICT Practice

Aside from having written policies, organisations have to make the policy work. This is attained through ICT practices. ICT practices are the actual application and realization of eco-sustainability considerations in ICT infrastructure sourcing, operations and disposal [2]. Molla and Cooper (2009) observe that it is possible for an organization to have policies in place and yet not have the same actualised. Green ICT Practice refers to the extent an ICT organization has translated its Green ICT concerns and policies into actions along the ICT activity value chain [1]. This is the actual application and realization of eco-sustainability considerations in ICT infrastructure sourcing, operation and disposal [9]. In particular, the sourcing and ICT infrastructure design and energy consumption audit and monitoring are important sub-components of Green ICT practice are components to pay attention to. The main areas for assessment of practice readiness according to Cooper and Molla (2009) are IT sourcing, IT operations and services and IT end of life management.

Policies can be made to work using several approaches such as setting green ICT goals to achieve, performing regular monitoring using the set matrix, having ICT personnel to champion green ICT, auditing the activities undertaking among other tasks. Green ICT sourcing practice readiness entails evaluating the environmental behaviour of suppliers, advocating of green technologies, and shortening of ICT infrastructure refresh periods.

Green ICT operation practices involve people, clients, and servers among other ICT infrastructure [1]. It deals with issues related to clearly articulate green guidelines for buying ICT equipment and services and adoption of environmental friendly purchases being undertaken [10]. These may be determined through assessing how the organisations operate existing ICT systems in an energy efficient manner, audits the power efficiency of existing ICT systems and technologies, enforces personal computer power management, analyses ICT's energy bill separately from overall corporate bill, implements ICT projects to monitor the organisation's carbon footprint, engages with a professional service provider regarding green ICT, and retires energy inefficient systems.

According to Molla & Cooper (2009:15), green ICT end of life management practice is the "compliance of IT equipment/machinery manufactures, users, and sellers in green IT end of life management." It has got to do with establishing whether ICT infrastructure as well as its packaging is reusable especially if their crushing or burning could harm the environment [11]. This may be actualised through recycles consumable equipment, disposes ICT equipment by returning it to suppliers, and disposes of ICT equipment in environmentally friendly manner and employing energy efficient coding. The principle behind energy efficient coding is to save power by getting software to make less use of the hardware, rather than continuing to run the same code on hardware that uses less power.

D. Green ICT Technology

Green ICT Technology refers to technologies and information systems for (a) reducing the energy consumption of powering and cooling corporate IT assets (such as data centers) (b) optimizing the energy efficiency of the IT technical infrastructure (c) reducing IT induced greenhouse gas emissions (d) supplanting carbon emitting business practices and (e) analyzing a business's total environmental footprint" [2]; [4]. From this perspective it may be said to reflect the extent to which the organizations acquire and build a more environmentally effective ICT infrastructure. This involves technologies and information systems for reducing the energy consumption of powering and cooling of ICT assets; optimizing the energy efficiency of the ICT technical infrastructure; reducing ICT induced greenhouse gas emissions; supplanting carbon emitting business practices; and analyzing a business's total environmental footprint [12]; [13]. The major question is "how to organisations use their ICT infrastructure?"

According to Cooper & Molla (2010), green ICT technology readiness may be measured by assessing the extent to which an organisation has green business infrastructure and green power sources, development of green ICT standards across the enterprise, server consolidation and virtualization, extent that applications and technologies are retired for greener technologies and extent of solutions development to support enterprise wide green initiatives.

As organisations institute their technology infrastructure they can employ a number of green ICT infrastructure technologies. With the continuous demand for ICT use in the developing economies as they strive for an e-economy there is an increased utilisation of servers. At the same time there is increased demand for servers following increased demand for networking, storage, speed, computation, backups and recovery which result in increased demand for energy and release of green houses gases [14]. However as they do this about 30% of the servers are dead just consuming power without being utilised therefore increasing costs of conducting business [15]. Green ICT utilisation comes in handy to reducing costs and green house gases emission.

E. Green ICT Governance

Green ICT Governance is the operating model that defines the administration of Green ICT initiatives and is closely related to the policy dimension [9]. It refers to the ICT management capability to put in place environmental criteria and frameworks to guide the sourcing, use, and disposal of the ICT technical infrastructure and activities of ICT personnel [2]. Effective ICT governance is the most important predicator of the value of an organisation generates from ICT [16]. According to Weill and Ross (2005), it is the practice that allocates decision rights and establishes the accountability framework for ICT decisions. And according to Schmidt & Kolbe (2011) therefore green ICT governance specifies the decision rights and accountability framework to encourage environmentally desirable behaviour in the sourcing, use and disposal of ICT. It is needed to establish clear roles, responsibilities, accountability and control of Green IT initiatives. For this to be achieved the roles responsibilities, accountability and control for green ICT initiatives need to be clearly established [4].

Well established responsibility structures on the matter may thus be a very good indicator of the organisations green ICT preparedness. It also implies allocating a budget for actualizing green ICT and putting in place metrics for establishing the impacts of green ICT initiatives [6]. In creating such structures the human personnel plays a major role. It has to provide sound management decision to understand impacts, prioritize actions and manages the enterprise's responses as required to implement successful Green ICT initiatives [4]. It is ICT technical human personnel that align ICT with the organisation goals. But this has to be done in line with what the managerial human personnel has set.

No single governance approach may be applied across organisations in line with green ICT [17]. This is entirely because specific external and internal factors of a given organisation, derivation from organisational settings, regulatory markets, socio-cultural, ecological, and technological environments are unique to each organisation [18] and often does influence the organisation's actions and this are directed by the human infrastructure. How the human personnel perceives the importance or uncertainty of green ICT determine the level of implementation of green ICT in the given organisation [19]. Every organisation attempts to encourage different behaviours [17] majorly dependents upon the managerial human personnel equipment with the necessary green ICT skills, a fact the model fails to consider. Without paying special attention to the managerial infrastructure therefore implies minimal chances of green ICT being successfully applied. Governance also involves change management, knowledge management, strategic planning and alignment and policies [20]; [21].

For governance to be actualize green ICT, it has to have clear structures in place where the role for coordinating green ICT initiatives are defined and the CEO plays a leading role in green ICT initiatives. It provides for responsibilities are clearly defined within each green ICT initiative which necessitates establishment of metrics for assessing the impact of green ICT initiatives and setting targets to reduce the organizations carbon footprint. Governance would realise much where ICT department is responsible for its own electricity bill.

III. METHOD

The study surveyed four cases that included a leading sugar manufacturing factory in the country that has invested into intensive application of ICT in virtually all its operations, a university that offers training programs in Information technology starting from certificate all the way to doctoral level, a communications commission involved in the regulatory tasks of information and communications technology in the country and senior government officers. From the four cases respondents were selected based on the involvement in ICT use. They consisted of top management, ICT technical uses, postgraduate (MSc and PhD) level students and senior government officers. The postgraduate students consists of a combination of information technology lecturers, ICT managers and ICT technical works in various companies and government departments in the country.

The study used a combination of questionnaire, interview and observation to collect data. This was complimented with secondary data from literature. The questionnaire used a seven scale for responses ranging from strongly disagree taking the value of 1 and being the lower most while strongly agree was the highest with the value 7. The neutral value had been assigned a value of 4. The respondents ranked their level of agreement with provided statements on the scale of seven (1. Strongly disagree 2. Disagree 3. Fairly disagree 4. Neutral 5. Fairly agree 6. Agree 7. Strongly agree).

IV. Findings of the Study

Green ICT readiness was assessed from five dimensions namely attitude, policy, practice, governance and technology maturity levels which is presented in this section (5.1). The relationship between the five dimensions of green ICT readiness and the ICT personnel characteristics in the study presented as demographics is also presented in this section (5.2).

A. Green ICT readiness in Kenya

The study found that the green ICT attitude level of organisations was very low in Kenya as compared to other countries such as the USA, New Zealand and America. For example, the study found the pervasiveness of green ICT within Kenya as at 2012 in comparison to other countries in the world such as Australia, New Zealand and USA as at 2009 and Indonesia as at early 2012 to be what is presented by Figure 2.

1) Green ICT readiness in Kenya

The study found that the green ICT attitude level of organisations was very low in Kenya as compared to other countries such as the USA, New Zealand and America. For example, the study found the pervasiveness of green ICT within Kenya as at 2012 in comparison to other countries in the world such as Australia, New Zealand and USA as at 2009 and Indonesia as at early 2012 to be what is presented by Figure 2

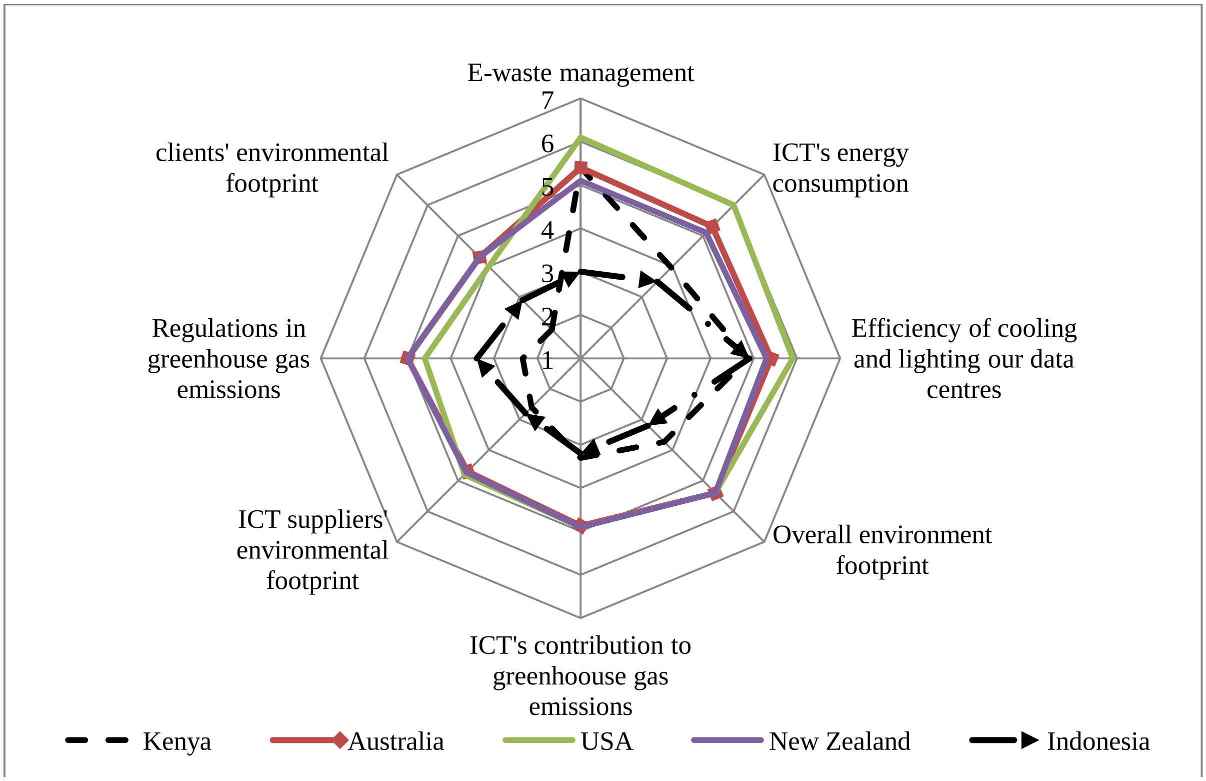


Figure2: Pervasiveness green ICT awareness an international perspective

Sources: Molla et al, 2009; Mariani & Imam, 2012; Field data 2012

From the findings presented in Figure 2, it is observable that the pervasiveness of green ICT awareness in Kenya is relative same as that of Indonesia except for e-waste concern. Kenya appears to be at the same level with Australia and New Zealand. Otherwise on the rest of the awareness aspects, Kenya is fairly below where Australia, USA and New Zealand were in 2009. There is need therefore deliberate effort to raise the awareness levels and have green ICT implemented. As the country works towards sustainability in its vision 2030, there is more to be done with regard to technology use than adopting it. It will be prudent to realize that excessive use of technology is not the way for achieving economical and technological development but the use of such technologies should be coupled with sound planning to ensure minimal adverse effects on the natural environment [23]. Sustainability is critical in ICT use hence the need to implement green ICT.

The Policy maturity level for Kenya in comparison to the USA, New Zealand, Australia and Indonesia is presented in Figure 3.

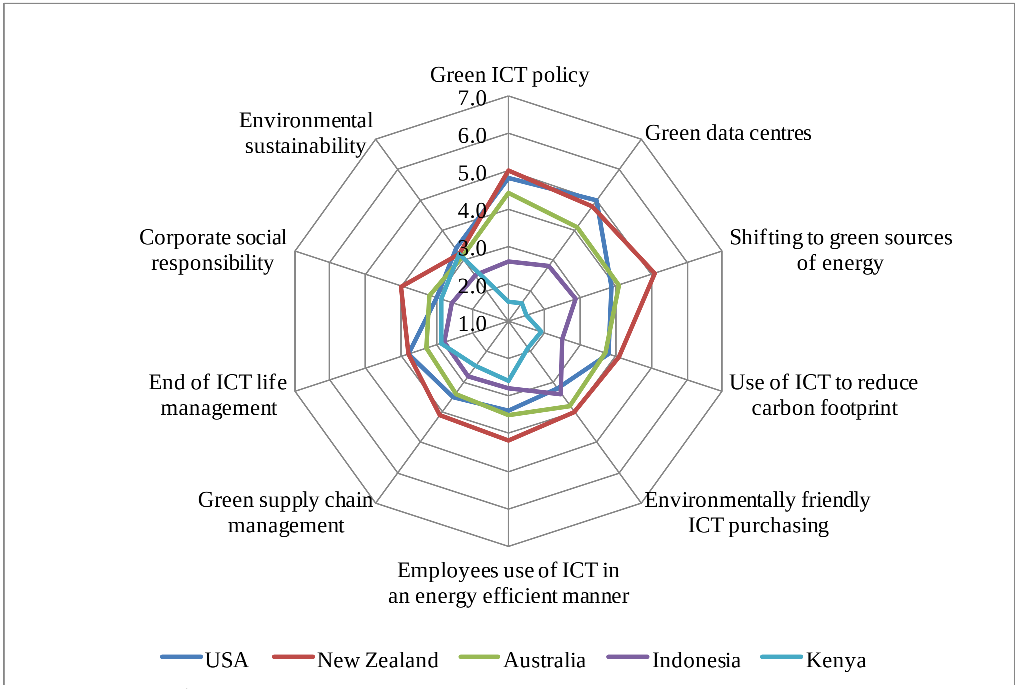


Figure 3: A comparison of countries green ICT policy maturity levels
Sources: Molla et al, 2009; Mariani & Imam, 2012; Field data 2012

From Fig. 3, it is clear that the green ICT policy maturity levels for developing economies (Kenya and Indonesia) are lower than those of the middle and developed economies (USA, Australia and New Zealand). The curves of Kenya and Indonesia are fairly closer to the value 1 being the lowermost scale point.

The practice maturity level was assessed through the sourcing maturity level and the operations maturity level. The findings of the sourcing maturity are presented in figure 4.

Figure 4 clearly shows that the sourcing maturity of Kenya and Indonesia is much lower as compared to USA, New Zealand and Australia. The curves of Kenya and Indonesia are inside those of the other countries. The operations maturity level findings are presented in Figure 5.

As can be seen from Figure 5, the operations maturity of Kenya and Indonesia are much lower as compared to USA, New Zealand and Australia. The curves of Kenya and Indonesia are inside those of the other countries. On the overall therefore Kenya’s and Indonesia’s practice maturity level is lower than that of the USA, New Zealand and Australia.

The technology maturity level was assessed via green infrastructure technologies and green data centre physical infrastructure maturity level. The findings are presented in Figure 6 and Figure 7.

On the dimension of governance, the findings of the study are presented in Figure 8.

Figure 6 shows that the maturity level of green ICT infrastructure technologies is low for Kenya and Indonesia as compared to the USA, Australia and New Zealand. On the other hand Figure 7 presents the data centre physical infrastructure maturity findings

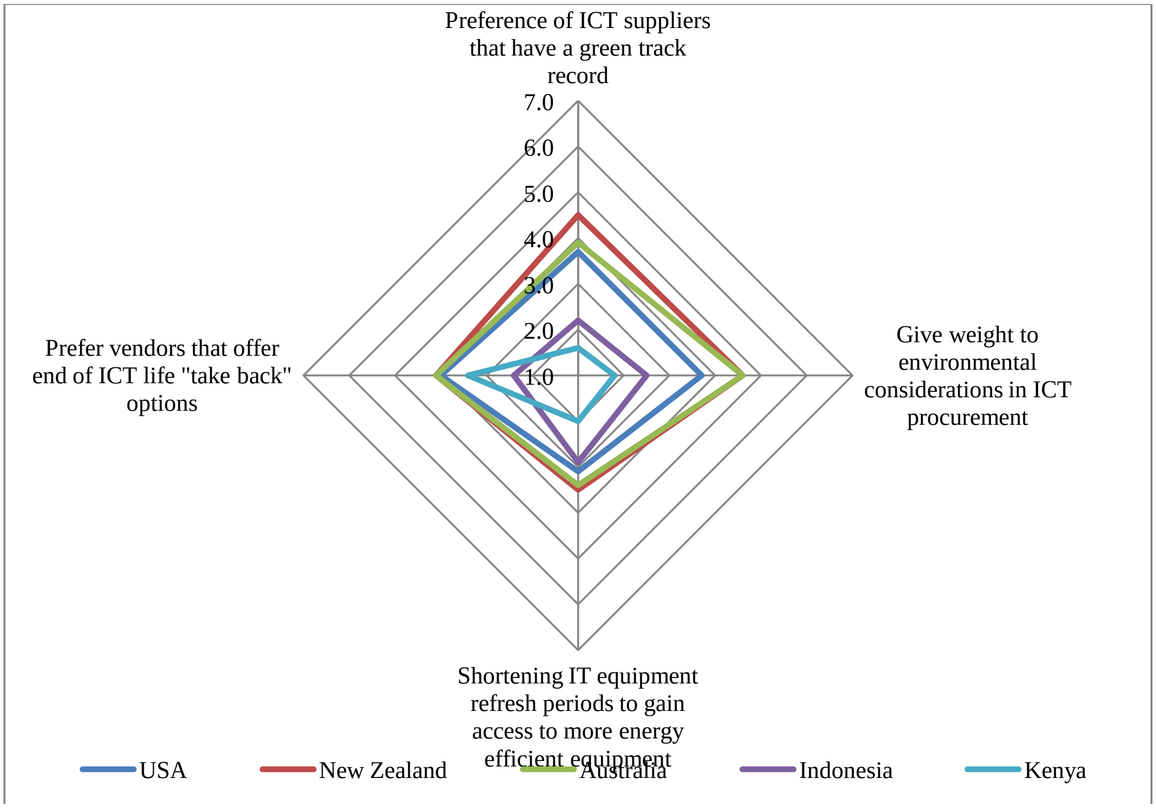


Figure 4: Countries comparison of green ICT sourcing maturity level
Sources: Molla et al, 2009; Mariani & Imam, 2012; Field data 2012

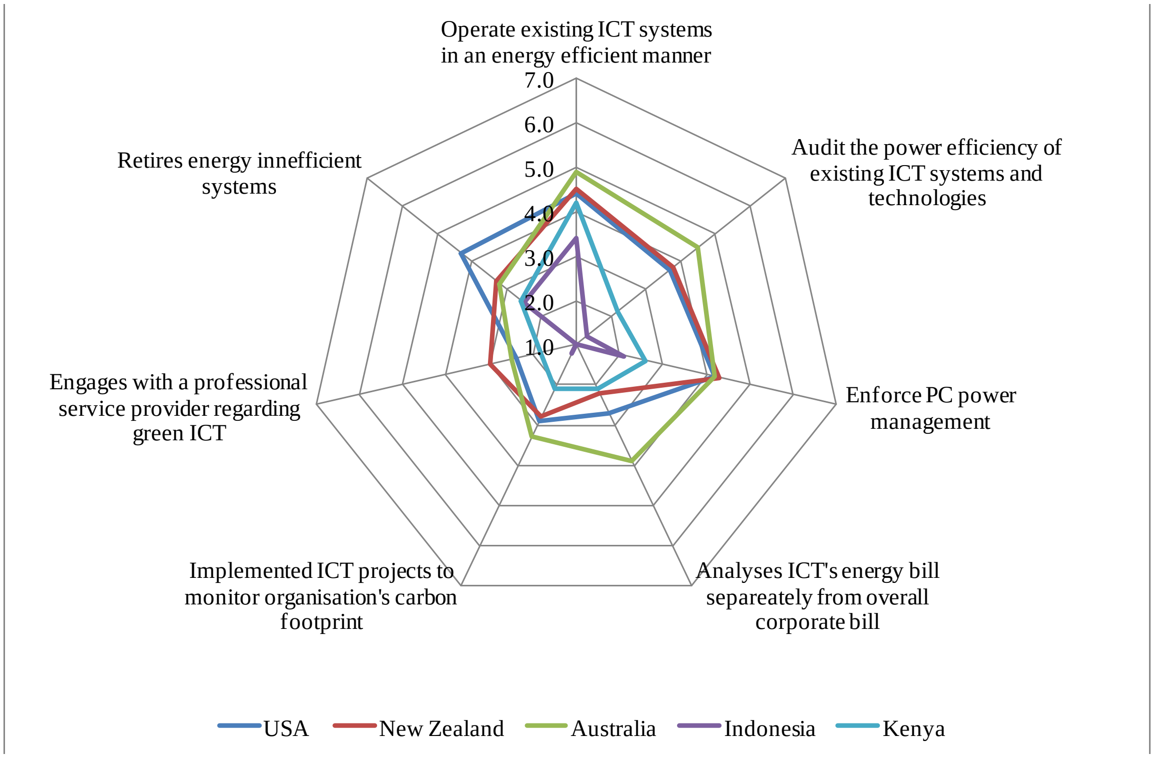


Figure 5: Countries comparison of green ICT operations maturity level
Sources: Molla et al, 2009; Mariani & Imam, 2012; Field data 2012

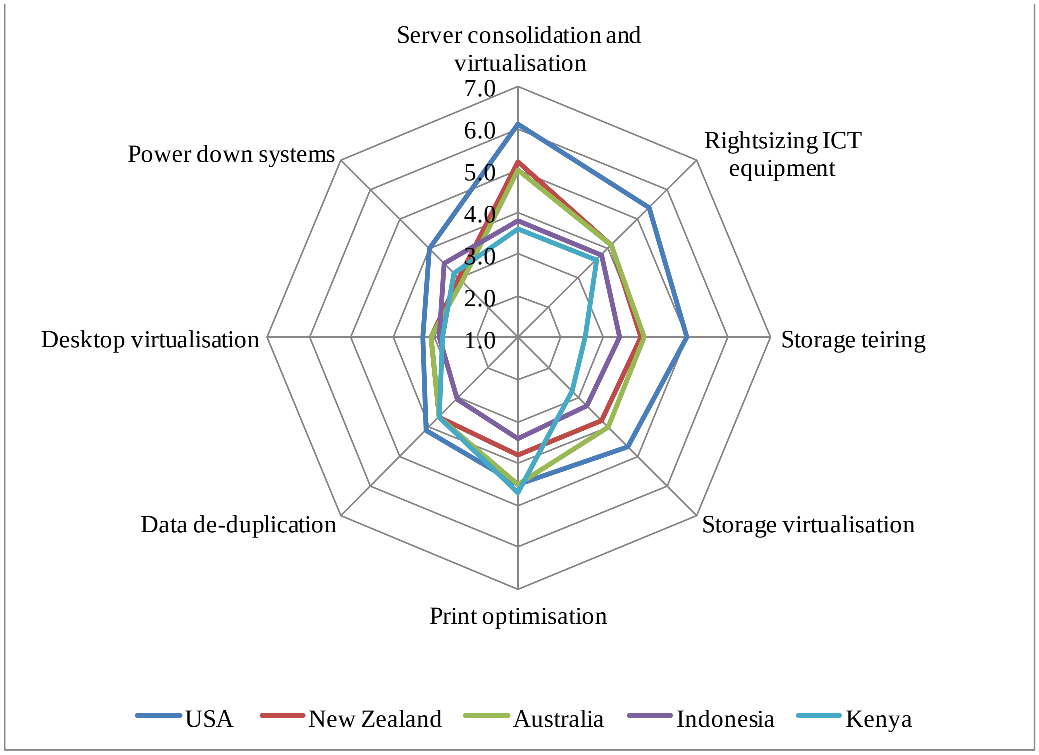


Figure 6: Countries comparison of green ICT infrastructure technologies maturity level
Sources: Molla et al, 2009; Mariani & Imam, 2012; Field data 2012

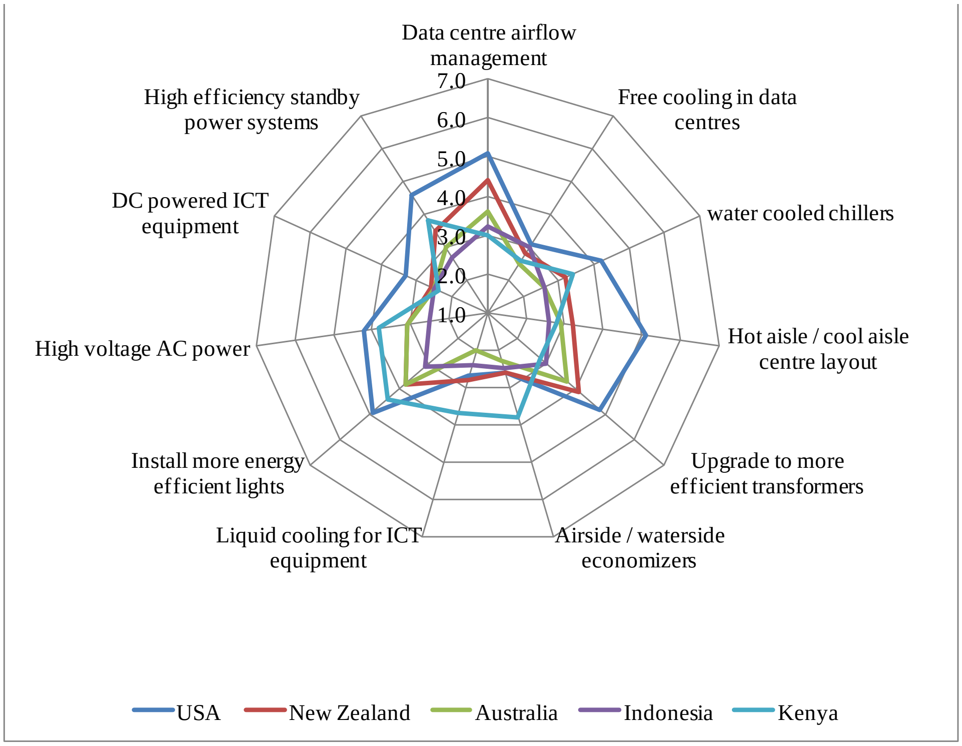


Figure 7: Countries comparison of green ICT data centre physical infrastructure maturity level
Sources: Molla et al, 2009; Mariani & Imam, 2012; Field data 2012

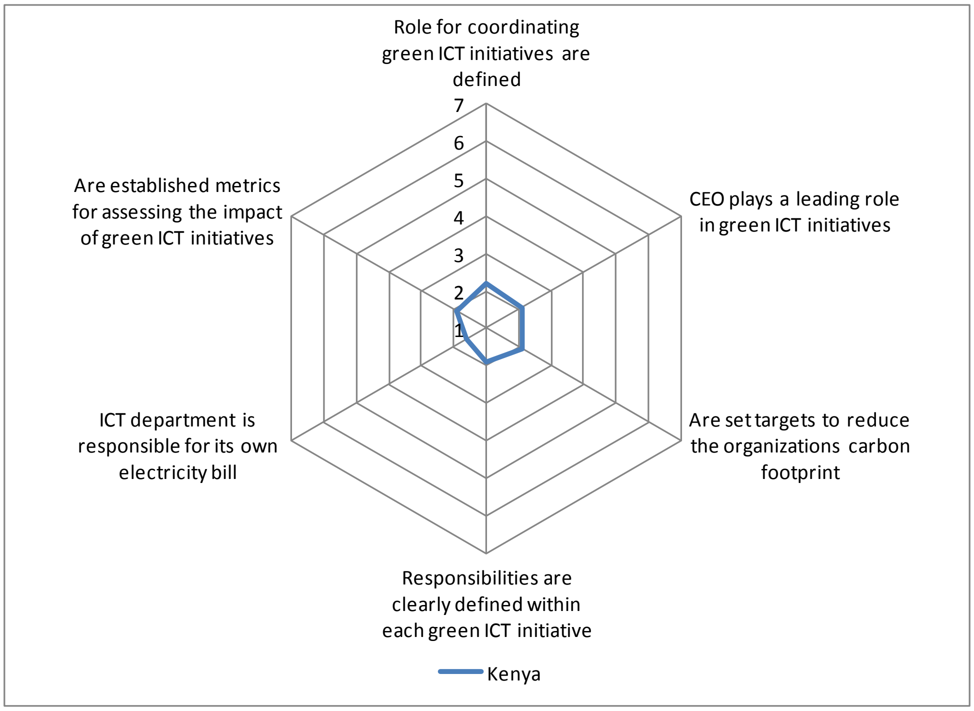


Figure 8: Green ICT governance maturity level in Kenya
Sources: Field data 2012

The green ICT governance level in Kenya is very low. The highest value on all aspects assessed is just slightly above 2 on a scale of 7. This clearly indicates that governance structures are fairly not in place or green ICT is yet to be given the attention it deserves.

The G-readiness level being low in all dimensions led to attempt to answer the question why it was lower in the developing economies as compared to middle and developed economies. This led establishing the relationships that existed between the personnel characteristics that were considered within the demographics of the study and the different dimensions of green ICT readiness.

B. Relationships between demographics and green ICT readiness dimensions

This section gives an overview of the findings of the ICT personnel characteristics who were involved in the study in section 5.2.1 and their relationship with the five dimensions of green ICT readiness following Molla et al model of G-readiness in section 5.2.2.

1) Demographics

The respondents for the study came from different sector and gender therefore had some basic characteristics that are explained in this section. The characteristics include age, gender, sector of work, occupation, academic qualifications and technical qualifications.

Figure 9 gives the findings on respondent's age.

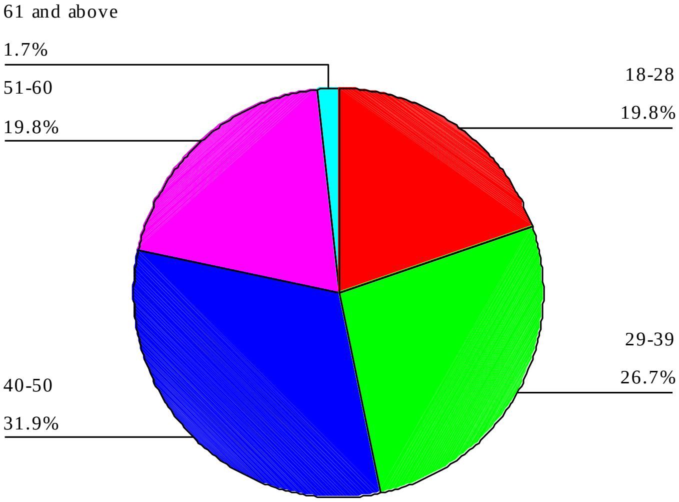


Figure 9: Respondents distribution based on age

From Figure 9 only 1.7% of the respondents were 61 years and above in age while 19.8% were of age between 51 and 60. In the age bracket of 40 to 50 there were 31.9% of the respondents with 26.7% being of age between 29 to 39 years. Only 19.8% of the respondents were of age between 18 to 28 years. Most of the respondents are of age between 29 to 50 years constituting 57.5%. The age set in employment therefore is equally distributed across the board however there are very few employees with over 50 years working within the ICT area.

This could be probably due to training in the field not having been there in the country in the earlier years. ICT is a young field in the developing nations like Kenya.

These respondents of the various ages were distributed across different occupation levels within their organisations. Figure 10 presents the findings of distribution of occupational levels amongst the respondents.

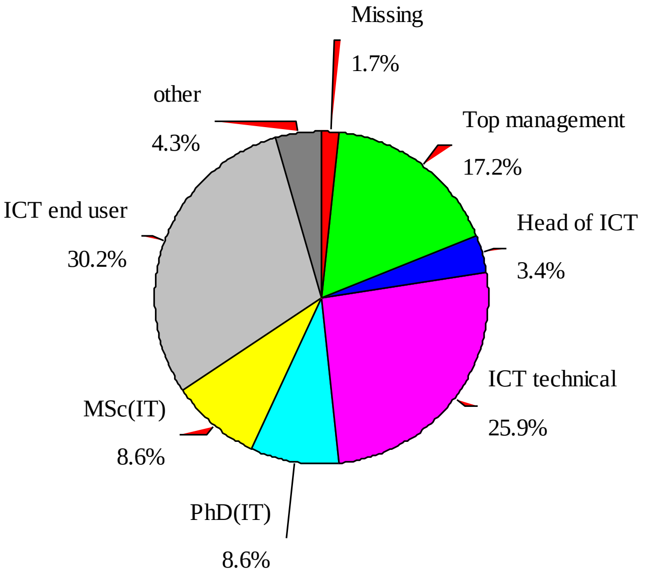


Figure 10: Respondents distribution based on their occupation

As can be seen from Figure 10, 1.7% the respondents did not indicate the occupation. Top management respondents constituted 17.2%. Top management play the critical of directing the organisations focus. The heads of ICT sections within the organisations made up 3.4% only. The ICT technical personnel constituted 25.9% of the respondents. The heads of ICT sections and the technical personnel are the key personnel in implementing ICT innovations. The end users who actual utilise or actualise the implemented technologies constituted 30.2%. The MSc (IT) and PhD (IT) students made up 17.2%. They were also composed of a selection of ICT professionals from different sections some of who ICT managers, lecturers and consultants. Since end users form the majority of the ICT personnel, there is need to have them to be made highly aware and equipped with green ICT skills if the implementation is to be enhanced.

Despite the different professions that were held by respondents, they were all either male or female with the gender findings being presented in Figure 11.

From Fig. 11, only 31% of the respondents were female with the rest, 69%, being male. Though the percentage of the female respondents is low, it’s worth noting that they are above one third. As compared to others sectors, the ladies seem to be doing fairly better in the ICT sector. However there is need to improve further the number of appropriately qualified ladies employed within the sector.

The respondents had diverse academic training levels as presented in Figure 12.

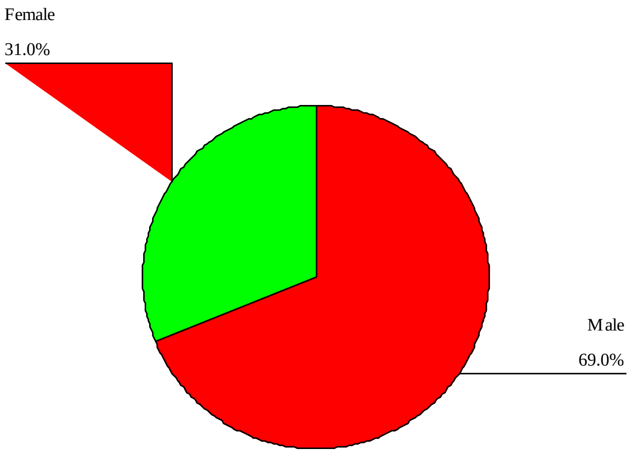


Figure 11: Respondents distribution based on gender

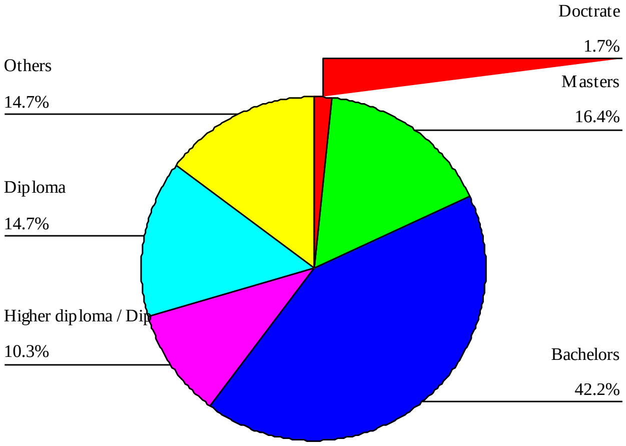


Figure 12: Respondents distribution based on the highest academic qualification

According to data presented in Fig. 12, only 1.7% of the respondents had a Doctorate degree, while 16.4% had Masters Degree. Forty two point two percent had Bachelors degree, 10.3% having a higher Diploma, 14.7% held a Diploma and 14.7% of the respondents had other academic qualifications. There are fairly low numbers of personnel in general trained at Masters and Doctorate level yet this are the cadres that lead in research undertakings.

When it came to relevant qualification in terms of ICT, their distribution of the highest technical qualification is presented in Figure 13.

From Figure 13, it is clear that only 13.8% of the respondents had Masters Degree in ICT related fields, 22.4% had Bachelors degree, 6.9% had higher Diploma, 6% had Diploma while the rest held other qualifications. An interview with a few of those who had other ICT related qualifications revealed that most respondents were holders of computer literacy skills (Basics / packages) certificates. The fact that there was hardly anyone with a doctorate and only 13.8% of masters degree holders suggests that there are few skilled qualified personnel in the ICT area. The lack is likely to have an impact on consultancy and research undertakings from within the country and developing nations on how best to apply green ICT in the local environment. The limited consultancy service in the area may likely also limit the awareness levels about the technology.

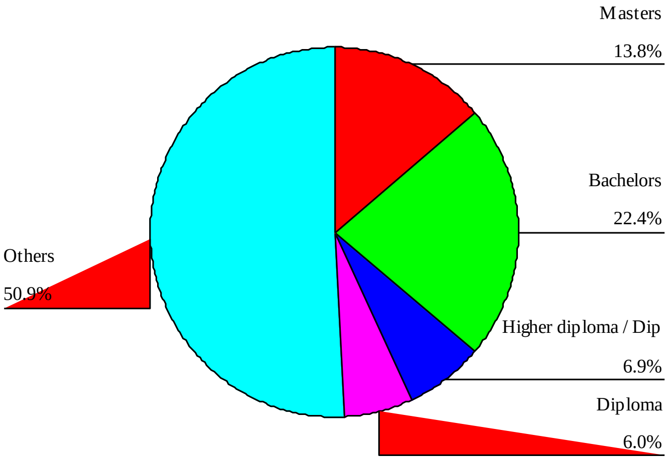


Figure 13: Respondents distribution based on technical ICT qualification

2) Spearman's rho coefficients of relationships between demographics and dimensions of G-readiness model

The first to be analysed was that of demographics with the aspects of pervasiveness. The findings are presented in table 1.

From the evaluations obtained in table 1, it can be observed that the age of respondents which may be considered to be directly related to ones experience has a significant relationship with the understanding of green ICT relevance awareness factors to an organisation's business. Based on spearman's rho coefficient values to the age of respondents, there is a moderate direct relationship at 0.01 (1-tailed) significance level with the understanding on reducing of costs of powering ICT infrastructure (0.284), purchasing more environmentally friendly ICT technology (0.281), use of ICT to minimise carbon emitting business practices (0.349) and improving energy efficiency of data centres (0.224). At the same time it has a weak indirect relationship with discarding ICT items in an environmentally friendly manner of coefficient 0.180 at a significant level of 0.05 (1-tailed). It is also apparent from the evaluations obtained in table 1 that there is no significant relationship with regard to gender and the green ICT awareness variables. Therefore the gender of the respondents has no influence on the awareness level of a given person and hence the organisations.

The highest general academic qualification seems to have a significant relationship with the green ICT awareness. From the spearman's rho coefficient values obtained the highest general academic qualification of respondents has a moderate indirect relationship at 0.01 (1-tailed) significance level with the use of ICT to minimise carbon emitting business practices (0.225), improving energy efficiency of data centres (0.256), and reducing the costs of running data centres (0.330). At the same time it has a direct weak relationship with reducing ICT's contribution to green house gas emissions (0.192) and complying with green regulatory requirements (0.173).

TABLE 1: SPEARMAN'S RHO CORRELATION COEFFICIENTS OF AWARENESS PERVASIVENESS AND ICT PERSONNEL VARIABLES

		Reducing the costs of powering ICT infrastructure	Purchasing more environmentally friendly ICT technology	Use of ICT to minimise carbon emitting business practices	Discarding ICT items in an environmentally friendly manner	Improving energy efficiency of data centres	Reducing the costs of running data centres	Reducing ICT's contribution to green house gas emissions	complying with green regulatory requirements
Age of respondent	Correlation Coefficient	.284 (**)	.281 (**)	.349 (**)	-.180 (*)	.224 (**)	.152	-.039	.098
	Sig. (1-tailed)	.001	.001	.000	.026	.008	.052	.338	.152
Respondent's gender	Correlation Coefficient	-.102	.002	.017	.056	-.152	-.027	.104	-.043
	Sig. (1-tailed)	.138	.490	.430	.275	.051	.385	.133	.326
Highest acad. qualifications	Correlation Coefficient	.025	.070	-.225 (**)	-.137	-.256 (**)	-.330 (**)	.192 (*)	.173 (*)
	Sig. (1-tailed)	.395	.226	.007	.072	.003	.000	.019	.035
Highest ICT qualification	Correlation Coefficient	.236 (**)	.159 (*)	.117	-.132	-.043	-.114	.163 (*)	.244 (**)
	Sig. (1-tailed)	.005	.044	.106	.079	.323	.111	.040	.005

* Correlation is significant at the 0.05 level (1-tailed).
** Correlation is significant at the 0.01 level (1-tailed).

Lastly, there is a significant relationship between the highest ICT qualification and green ICT awareness factors. According to spearman’s rho coefficient values obtained the highest technical academic qualification of respondents has a indirect relationship at 0.01 (1-tailed) significance level with reducing the costs of powering ICT infrastructure that is moderate (0.236), purchasing more environmentally friendly ICT technology that is weak (0.159), reducing ICT's contribution to green house gas emissions that is weak (0.163) and complying with green regulatory requirements that is moderate at 0.244.

In order to establish how the ICT personnel related variables were related with the green ICT drivers, spearman’s rho coefficients were evaluated and the results are presented in table 2.

Table 2: Spearman’s coefficient correlation between the ICT personnel variables and green ICT drivers

		Age of respondent	Respondent's gender	Highest academic qualifications	Highest ICT qualification
E-waste management	Coefficient	.146	-.157 (*)	-.375 (**)	-.123
	Sig. (1-tailed)	.061	.049	.000	.098
Efficiency of cooling and lighting data centre	Coefficient	.179 (*)	-.091	.129	.232 (**)
	Sig. (1-tailed)	.030	.170	.089	.007
Overall environmental footprint	Coefficient	.301 (**)	-.076	-.204 (*)	-.027
	Sig. (1-tailed)	.001	.210	.015	.390
ICT's contribution to green house gas emissions	Coefficient	.207 (*)	.100	-.048	.187 (*)
	Sig. (1-tailed)	.014	.146	.307	.023
ICT suppliers' environmental footprint	Coefficient	.118	-.040	.324 (**)	.276 (**)
	Sig. (1-tailed)	.110	.337	.000	.002
Regulations in green house gas emissions	Coefficient	.265 (**)	.216 (**)	.181 (*)	.400 (**)
	Sig. (1-tailed)	.002	.010	.026	.000

* Correlation is significant at the 0.05 level (1-tailed).
** Correlation is significant at the 0.01 level (1-tailed).

From the evaluations obtained in table 2, it can be observed that the age of respondents has a significant relationship with the green ICT drivers of an organisation’s business. Based on spearman’s rho coefficient values to the age of respondents, there is a weak direct relationship at 0.05 (1-tailed) significance level with efficiency of cooling and lighting data centre at 0.179, a moderate direct relationship at 0.01 (1-tailed) significance level with overall environmental footprint at 0.301of 0.207, and a moderate direct relationship of 0.265 at 0.05 (1-tailed) significance level with ICT's contribution to green house gas emissions.

It is also apparent from the evaluations obtained in table 2 that there are only two items of the eight where there is a significant relationship with regard to gender and the green ICT drivers. The only direct significant relationship between gender and green ICT drivers is has moderate relationship found on the regulations in green house gas emissions with spearman’s rho coefficient of 0.216 at 0.01, (1-tailed) significance. The other which is an indirect significant relationship between gender and green ICT drivers is a weak relationship found with e-waste management that has a spearman’s rho coefficient of -0.157 at 0.05, (1-tailed) significance. Therefore the gender of the respondents has no major influence on green ICT drivers likely to influence a given person and hence the organisations to implement green ICT. The general highest academic qualifications seem to have a significant relationship with ICT drivers. According to the results in table 1 it can be observed that there exists an indirect moderate relationship with E-waste management (-0.375) and overall environmental footprint (-0.204). There is a direct relationship also with ICT suppliers' environmental footprint that is moderate (0.324) and a weak relationship with regulations in green house gas emissions (0.181).

The technical qualifications seem to have a significant relationship with green ICT drivers. With exception of ICT's contribution to green house gas emissions where the relationship is a direct weak one at 0.187, (1-tailed), significance level of 0.05, the rest have a moderate direct relationship with a 0.01 significance level. Efficiency of cooling and lighting data centre has spearman’s rho value of 0.232, ICT suppliers' environmental footprint has spearman’s rho value of 0.276 and regulations in green house gas emissions has spearman’s rho value of 0.400. The findings of significant relationships between the demographics and the policy and practice maturity dimensions are presented in table 3.

TABLE 3: SPEARMAN'S RHO COEFFICIENT CORRELATIONS BETWEEN ICT PERSONNEL VARIABLES AND GREEN ICT POLICY AND PRACTICE MATURITY VARIABLES

		Age of respondent	Respondent's gender	Highest academic qualifications	Highest ICT qualification
Policy maturity					
Green ICT	Correlation Coefficient	.148	.048	.319 (**)	.341 (**)
	Sig. (2-tailed)	.117	.615	.001	.000
Environmentally friendly purchasing	Correlation Coefficient	-.060	.207(*)	.135	.101
	Sig. (2-tailed)	.533	.029	.158	.290
Corporate social responsibility	Correlation Coefficient	-.039	.067	.269 (**)	.168
	Sig. (2-tailed)	.678	.475	.004	.072
ICT practice maturity					
	ICT sourcing				
Prefers ICT suppliers that have a green track record	Correlation Coefficient	-.008	.191(*)	-.132	-.098
	Sig. (2-tailed)	.929	.042	.162	.301
Is shortening ICT equipment refresh periods to gain access to more efficient energy equipment	Correlation Coefficient	.095	.066	.181	.195(*)
	Sig. (2-tailed)	.310	.483	.052	.036
ICT operations practice					
Enforce personal computer power management	Correlation Coefficient	.093	-.015	.103	.201(*)
	Sig. (2-tailed)	.329	.875	.278	.033
Implements ICT projects to monitor the organisation's carbon footprint	Correlation Coefficient	.195(*)	-.082	-.148	-.045
	Sig. (2-tailed)	.036	.383	.112	.634
End life management					
Disposes of ICT equipment in environmentally friendly manner	Correlation Coefficient	.040	.006	-.068	-.315 (**)
	Sig. (2-tailed)	.670	.947	.471	.001

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

From table 3, implementation of ICT projects to monitor the organisation's carbon footprint had a significant weak positive (0.195) relationship with the age of the respondents at a 2-tailed significance level of 0.05. The respondents gender had a significant moderate positive relationship of 0.207 on environmentally friendly purchasing and 0.191 on prefers ICT suppliers that have a green track record at a 2-tailed significance level of 0.05.

For the highest academic qualification there was a significant moderate positive relationship of 0.319 and 0.269 with green ICT and corporate social responsibility respectively at 0.01 in 2-tailed significance level. Finally with regard to the highest ICT training had a positive moderate relationship of 0.341 at 0.01 in 2-tailed significance level, shortening ICT equipment refresh periods to gain access to more efficient energy equipment had a positive weak relationship of 0.0.195 at 0.05 in 2-tailed significance level, enforcement of personal computer power management had a had a positive weak relationship of 0.201 at 0.05 in 2-tailed significance level and finally the disposal of ICT equipment in environmentally friendly manner had a had a negative moderate relationship of 0.315 at 0.001 in 2-tailed significance level.

TABLE 4: SPEARMAN'S RHO COEFFICIENT CORRELATIONS BETWEEN ICT PERSONNEL VARIABLES AND GREEN ICT TECHNOLOGY AND GOVERNANCE MATURITY VARIABLES

		Age of respondent	Respondent's gender	Highest academic qualifications	Highest ICT qualification
Technology					
	ICT infrastructure				
Rightsizing ICT equipment	Correlation Coefficient	-.005	-.095	-.032	.214(*)
	Sig. (2-tailed)	.962	.320	.741	.023
Storage teiring	Correlation Coefficient	.081	-.060	.142	.260 (**)
	Sig. (2-tailed)	.387	.526	.130	.005
Storage virtualisation	Correlation Coefficient	.152	.118	.039	.185(*)
	Sig. (2-tailed)	.104	.207	.677	.047
Print optimization	Correlation Coefficient	-.107	.117	.263 (**)	-.047
	Sig. (2-tailed)	.258	.215	.005	.623
Power down systems	Correlation Coefficient	.123	-.169	-.215 (*)	-.059
	Sig. (2-tailed)	.190	.070	.021	.529
	Network infrastructure technologies				
DC powered ICT equipment	Coefficient	.033	.043	.268 (**)	.248 (**)
	Sig. (2-tailed)	.732	.650	.005	.009
High efficiency standby power system	Coefficient	.188(*)	-.062	.129	.234(*)
	Sig. (2-tailed)	.048	.516	.174	.013
Install energy efficient lights	Coefficient	.108	.045	.148	.203(*)
	Sig. (2-tailed)	.251	.634	.115	.030
ICT governance					
CEO plays a leading role in green ICT initiatives	Correlation Coefficient	.117	.082	.078	.299 (**)
	Sig. (2-tailed)	.217	.387	.409	.001
Targets are set to reduce the organisations carbon footprint	Correlation Coefficient	-.143	.040	.235(*)	.107
	Sig. (2-tailed)	.134	.673	.013	.261

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

From table 4, it can be observed that the respondent's age had only one significant weak positive relationship of 0.188 at a 2-tailed 0.05 significance level with the high efficiency standby power system. Respondents' gender had no significant relationship with any of the elements of the green ICT technology and ICT governance. The highest academic qualification had a positive moderate relationship of 0.0.263 at 0.01 in 2-tailed significance level, a negative weak relationship of 0.215 at 0.05 in 2-tailed significance level with print optimization, a had a positive moderate relationship of 0.268 at 0.01 in 2-tailed significance level with DC powered ICT equipment and had a positive weak relationship of 0.235 at 0.05 in 2-tailed significance level with targets being set to reduce the organisations carbon footprint. Finally the highest ICT technical qualification had a positive weak relationship of 0.214 at 0.05 in 2-tailed significance level with rightsizing ICT equipment, had a positive moderate relationship of 0.260 at 0.01 in 2-tailed significance level with storage teiring, had a positive weak relationship of 0.185 at 0.05 in 2-tailed significance level with storage virtualisation, had a positive moderate relationship of 0.248 at 0.01 in 2-tailed significance level with DC powered equipment, had a positive weak relationship of 0.234 at 0.05 in 2-tailed significance level with high efficiency standby power system, had a positive weak relationship of 0.203 at 0.05 in 2-tailed significance level with installation of energy efficient lights, and had a positive moderate relationship of 0.299 at 0.01 in 2-tailed significance level with the CEO playing a leading role in green ICT initiatives.

V. G-READINESS MODEL EXTENSION

From the analysis of relationships presented in IV B, it was found that the existed significant weak or moderate relationships at significance level of 0.01 or 0.05 between the Green ICT, environmentally friendly purchasing, corporate social responsibility, prefers ICT suppliers that have a green track record, shortening ICT equipment refresh periods to gain access to more efficient energy equipment, ICT operations practice, enforce personal computer power management, implements ICT projects to monitor the organisation's carbon footprint, end life management, and disposes of ICT equipment in environmentally friendly manner elements within the dimensions of ICT policy and practice. There was also significant weak or moderate relationship at significance level of 0.01 or 0.05 between rightsizing ICT equipment, storage teiring, storage virtualisation, print optimization, power down systems, DC powered ICT equipment, high efficiency standby power system, install energy efficient lights, CEO plays a leading role in green ICT initiatives and targets are set to reduce the organisations carbon footprint of the technology and ICT governance dimensions.

Based on the fact that the G-readiness levels of developing economies being lower than those of the middle and developed economies when measured on the Molla's G-readiness model and the many element's relationship between ICT personnel variables with the G-readiness dimensions, an extension of the model was made as provided for in Figure 14.

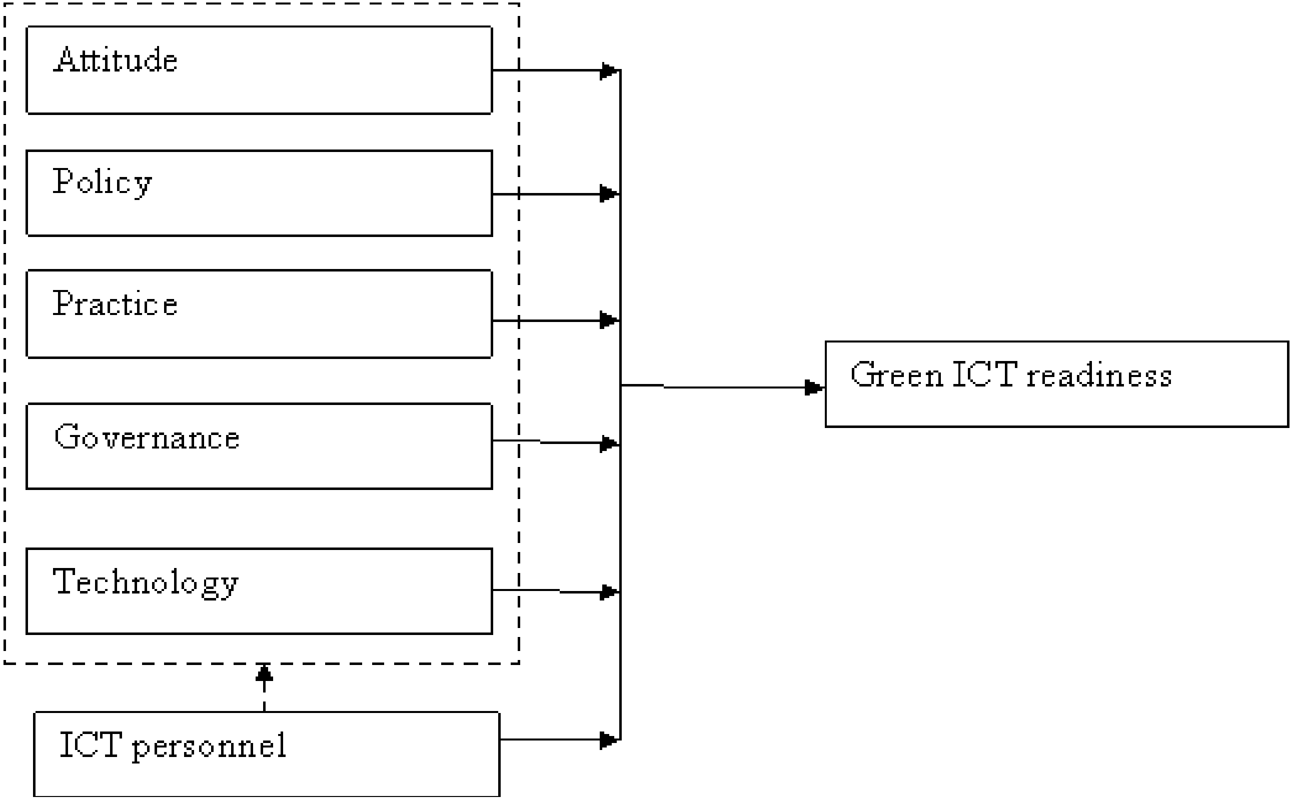


Figure 14: Extended Molla's green ICT readiness model for organisations in developing economies

In the extended version herein given, the study proposes that ICT personnel preparedness contributes directly to how an organisation gets ready in terms of green ICT readiness and yet has influence on the level at which the other five dimensions get to be ready as witnessed by the moderate significant relationships discussed in section 5.2.2. For this study the first four aspects within the ICT personnel to be considered were gender, experience (age), academic training and technical ICT (green ICT) training.

VI. CONCLUSION AND RECOMMENDATION

The study established that the level of green ICT readiness is lower in developing economies. It also established that there was a significa nt relationship between ICT personnel variables and green ICT preparedness dimensions. Based on the established significant relationships, it proposed an extended G-readiness model from the initial Molla's model for establishing the level of green ICT readiness for an organisation.

The study recommends further studies to be conducted to establish the full range of ICT personnel characteristics on green ICT readiness.

ACKNOWLEDGMENT

We would like to acknowledge the input by the Kibabii University College community and Prof. Leonard Wamocho during the research.

REFERENCES

[1] Molla A. & Cooper V. (2009). Green IT readiness: A framework and preliminary proof of concept. *Australasian journal of information systems*, 16(2): 5-23.

[2] Cooper V. A. & Molla A. (2012). Developing green IT capability: An absorptive capacity perspective. *Pacific asia Conference on Information Systems* (p. T5_148doc). www.pacis2012.org/files/papers/pacis2012_T5_cooper_148.doc.

[3] Wabwoba F., Wanyembi W. D. & Omuterema S. (2012). Barriers to green ICT implementation in Kenya. *International Journal of Science and Technology*. 2(12): 823-836

[4] Molla A., Cooper V. A., & Pittayachawan S. (2009). IT and eco-sustainability: Developing and validating a green IT readiness model. *Thirtieth International Conference on Information Systems*, (p. 17).

[5] Philipson G. (2010). *A Green ICT Framework: Understanding and Measuring Green ICT.* Australia: 2010 Connection Research Services Pty Ltd (ABN 47 092 657 513).

[6] Cooper V. & Molla A. (2010). *Conceptualizing Green IT Organzational Learning (GITOL). Green IT Working Paper Series. Paper No. 3/2010.* Melbourne: School of Business Information Technology and Logistics, RMIT University.

[7] Boughton, C. (2009). What is an ICT professional anyway. *Australasian journal of information systems*, 16(1):149-163.

[8] Rao P. & Holt D. (2005). Do green supply chains lead to competitiveness and economic performance? *International Journal of Operations & Production Management*, 25 (9): 898-916.

[9] Tenhunen, M. (2011). *Conceptualizing and Measuring Green IT Readiness in Finnish Companies. Application Area: Electronic Invoice.* Aalto University (Master's Thesis).

[10] [10] Velte T. J., Velte A. T. & Elsenpeter R. (2008). *Green IT: reduce your information system's environmental impact while adding to the bottom line.* New York: McGraw Hill.

[11] Alsever, J. (2008). *The 'green' way to dump electronic junk.* Retrieved February 12, 2012, from www.mnsbc.com/id/24163506/

[12] Chen A., Boudreau M., & Watson R. (2008). Information systems and ecological sustainability. *Journal of systems and information Technology, Sustainability and Information Systems* , 24 (3): 186-201.

[13] Elliot S.,& Binney D. (2008). Environmentally Sustainable ICT: Developing corporate capabilities and an industry relevant IS research agenda. *Proceedings of PACIS, 4-7 July Suzhuo, China.*

[14] Uddin M. and Rahman A. A. (2011). Virtualization implementation model for cost effective and efficient data centers. *International Journal of Advanced Computer Science and Applications*, 2 (1): 69-74.

[15] Mueen U. & Azizah A. R. (2012). Server consolidation: An approach to make data centers energy efficient and green. *International journal of science and engineering research*, 1(1): 1-7.

[16] [16] Weill P. & Ross J. W. (2004). *IT governance: how top performers manage IT decision rights for superior results.* Harvard Business Press.

[17] Schmidt N. & Kolbe L. M. (2011). Towards a contingency model for green IT Governance. *AMCIS 2011 Proceedings.*

[18] Jenkin T. A., Webster J. & McShane L. (2010). An agenda for "Green" information technology and systems research. *Information and organisation*.

[19] 1Schmidt N., Erek K., Kolbe L. M. & Zarnekow R. (2010). Predictors of green IT adoption: Implications from an Emprical Investigation. *AMCIS 2010 Proceedings.* available at: http://aisel.aisnet.org/amcis2010/367.

[20] Ogunyemi A. A. & Johnston A. K. (2012). Towards an organisational readiness framework for emerging readiness for technologies: An investigation of antecedents for South African organisations' readiness for server virtualisation. *The electronic Journal on information systems in developing countries* , 53(5):1-30.

[21] Maranto-Vargas D.and Rangel R. G. T . (2007). Development of internal resources and capabilities as sources of differentiation of SME under increased global competition: A field study in mexico. *Technicalogical forecasting and social change*, 70:90-99.

[22] [22] Mariani M. & Imam K. (2012). A Preliminary Study of Green IT Readiness in Indonesian Organizations. *Journal of Energy Technologies and Policy*, 2 (5): 1-10.

[23] [23] Olofsson A & Ohson S. (2006). General belief and environmental concerns: Transatlantic comparisons. *Environment and Behaviour*, 38: 768-790.

AUTHORS PROFILE

Mr. Franklin Wabwoba is a lecturer at Kibabii University College (Kenya) in the Department of Information Technology. He holds a Master of Science (Computer Applications) degree from Kenyatta University; Endorsement (Educational Management) from University of South Africa and Bachelor of Education (science: Mathematics and Computer Science) degree from Egerton University and is a PhD (Information Technology) candidate at Masinde Muliro University of Science and Technology. He has taught Computer Science and Information Technology courses for many years. He has ICT industrial experience having worked with Mumias Sugar Company. He has presented several papers in scientific conferences and has a number of publications in referred journals. He has a strong research interest in green ICT, the impact of ICT applications on the community and integration of ICT into education. He is a professional member of the Association for Computing Machinery (ACM).

Dr. Gregory Wabuke Wanyembi (PhD) obtained his doctorate in management of ICTs from Delft University of Technology, The Netherlands, in 2002. He has published widely in the area of management and strategic utilization of ICTs in developing countries. He is engaged in several research projects and supervises a number of postgraduate students at both masters and PhD drawn from universities in Kenya. He is also a reviewer of journal articles, and an external examiner at two universities. He is a former senior lecturer and chairman at the Department of Computer Science, Masinde Muliro University of Science and Technology, Kenya.

Dr. Stanley Omuterema Oluchiri (PhD) is a lecturer in Department of Disaster Management and Sustainable Development at Masinde Muliro University of Science and Technology. He held the position of centre coordinator of the Nairobi campus of Masinde Muliro University of Science and Technology for a number of years. He obtained his doctorate Moi University. He has published widely. He has wide experience having taught in secondary school, Eldoret Polytechnic, and Masinde Muliro University of Science and technology.

Mr. Kelvin Kabeti Omieno is an Assistant Lecturer in the Department of Computer Science, Masinde Muliro University of Science and Technology, Kenya. He holds MSc in IT and First Class Honors Degree in Computer Science from Masinde Muliro University of Science and Technology (Kenya). He is currently pursuing PhD in Information Technology of Strathmore University (Kenya). He has been involved in a number of research projects of ICTs and development, including computational grid project, and E-waste management in Kenya. Mr. Omieno has published widely in journals and conference proceedings in Information technology and ICTs for development. He is a professional member of the Association for Computing Machinery (ACM).

Evaluating English to Arabic Machine Translation Using BLEU

Mohammed N. Al-Kabi
Faculty of Sciences & IT
Zarqa University
Zarqa – Jordan

Taghreed M. Hailat, Emad M. Al-Shawakfa, and Izzat M. Alsmadi
Faculty of IT and CS
Yarmouk University,
Irbid 211-63, Jordan.

***Abstract*— This study aims to compare the effectiveness of two popular machine translation systems (Google Translate and Babylon machine translation system) used to translate English sentences into Arabic relative to the effectiveness of English to Arabic human translation. There are many automatic methods used to evaluate different machine translators, one of these methods; Bilingual Evaluation Understudy (BLEU) method, which was adopted and implemented to achieve the main goal of this study. BLEU method is based on the assumptions of automated measures that depend on matching machine translators' output to human reference translations; the higher the score, the closer the translation to the human translation will be. Well known English sayings in addition to manually collected sentences from different Internet web sites were used for evaluation purposes. The results of this study have showed that Google machine translation system is better than Babylon machine translation system in terms of precision of translation from English to Arabic.**

***Keywords*— component; Machine Translation; Arabic; Google Translator; Babylon Translator; BLEU**

I. Introduction

Machine translation is a task that involves the process of translating a source sentence from one language giving the meaning into another target language(s). Online machine translators rely on different approaches to translate from one natural language into another, these approaches are: Rule-based, Direct, Interlingua, Transfer, Statistical, Example-based, Knowledge-based, and Hybrid Machine Translation (MT).

The accuracy of any machine translator is usually evaluated by comparing the results to human judgments. One of the methods used to evaluate machine translation systems is called BiLingual Evaluation Understudy (BLEU) which was introduced in the study of Papineni, Roukos, Ward, and Zhu [1] and claimed to be language independent and highly correlated with human evaluation.

BLEU is based on a core idea to determine the quality of any machine translation system which is summarized by the closeness of the candidate output of the machine translation system to reference (professional human) translation of the same text.

The closeness of the candidate translation to the reference translation is determined by a modified n-gram precision which was proposed by Papineni, Roukos, Ward, and Zhu [1]. The modified n-gram precision is the main metric adopted by BLEU to distinguish between good and bad candidate translations, where this metric is based on counting the number of common words in the candidate translation and the reference translation, and then divides the number of common words by the total number of words in the candidate translation. The modified n-gram precision penalizes candidate sentences found shorter than their reference counter parts, also it penalize candidate sentences which have over generated correct word forms.

The US National Institute of Standards and Technology (NIST) have presented a new method called NIST; which represents an enhancement to BLEU. The NIST method is used to evaluate the effectiveness of a number of machine translation systems to translate from various natural languages into English. This method, and according to Doddington [2], tries to compute how informative a particular n-gram is, where a low frequency of a particular n-gram means yielding a higher weight, while a high frequency of a particular n-gram means yielding a lower weight.

Due to its rich and complex morphological features, Arabic has always been a challenge for machine translation. In addition, Arabic has different word forms and word orders which make it possible to express any sentence in different forms.

Furthermore, the existence of many dialects and the fact that the word order is not usually the same for source and target languages, this leads usually to the possibility of having more than one meaning for the same sentence according to Alqudsi, Omar, and Shaker [21]. English-to-Arabic machine translation has been a challenging research issue for many of the researchers in the field of Arabic Natural Language Processing.

Many attempts were made to perform or enhance machine translation of Arabic into other languages. Some of these attempts are the work of Al Dam, and Guessoum [3], Carpuat, Marton, and Habash [4], Adly and Al-Ansary [5], Salem, Hensman, and Nolan [6], and Riesa, Mohit, Knight, and Marcu [7].

To evaluate any translation system, one should use a proper corpus. As for Arabic, the authors could not find any standard corpus that could be used for evaluation purposes.

For this, we had to collect our data from different Internet websites representing two types of datasets; a set of well-known English sayings and a set of sentences that were translated manually by two human translators for judgment purposes. In this study, we have evaluated the effectiveness of two automatic machine translators that could be used for English-to-Arabic translation and vice versa. The used machine translators are Google machine translator and the Babylon machine translator.

This paper is organized as follows: section 2 presents the related work, section 3 presents the methodology followed in this research, section 4 presents the evaluation of machine translators under consideration, through the usage of a system designed and implemented by one of the authors. Section 5 presents the conclusion from this research, and last but not least section 6 discusses extensions of the this study and the future plans to improve it.

II. Literature Review

A number of studies were conducted to evaluate the translation quality using an automated tool. One of these researches was conducted by Nießen, Och, Leusch, and Ney [8]. In their study, they have presented the typical necessary requirements to build an effective tool for the evaluation of the accuracy of different machine translators. Word Error Rate (WER) and Subjective Sentence Error Rate (SSER) are discussed as two essential criteria to the quality of the outputs of machine translators. The authors have described their technique as fast, semiautomatic, convenient and consistent.

Precision, Recall, and F-measure are famous measures which are usually used to evaluate information retrieval systems (IRS) and search engines, however, in their study Melamed, Green, and Turian [9], have showed that these three measures can also be used to evaluate machine translators, and further showed that these three measures are highly correlated to these measures. In addition, the authors claimed that these measures are more reliable than Bilingual Evaluation Understudy (BLEU).

Usually, Machine translation evaluation methods are based on reference translations, but that is not always the case. So for example, Palmer [10] has introduced a user-centered method in his study to evaluate machine translation systems that is based on comparing the outputs of machine translation systems and then ranked, according to their quality, by expert users who have the necessary needed scientific and linguistic backgrounds to accomplish the ranking process. His study covers four Arabic-to-English and three Mandarin (simplified Chinese)-to-English machine translation systems.

Another method for evaluating machine translation systems was presented by Akiba et al. [11]. Their study was dedicated to evaluate machine translation (MT) systems that are subsystems of speech-to-speech MT (SSMT) systems. The researchers referred to the two drawbacks of using BLEU to evaluate SSMT, where the first drawback was related to the position based error assessment, while the second drawback was related to the tolerance to accept colloquial sentences. The new method presented in their paper was called “gRader based on Edit Distances (RED)”, which automatically computes the score related to the translated output of the machine translation system using a decision tree (DT). They have conducted a series of experiments which revealed; according to the authors, that their novel method RED is more accurate than BLEU method.

The BLEU method is characterized by the fact that it is language independent and not designed for a certain natural language. BLEU has a number of cons, therefore a number of researchers have attempted to enhance this important method. One of such attempts was conducted by Yang et al. [12]. They have used linguistic features of the evaluated sentences outputted from the machine translation systems in their enhancements. Those researchers have used multiple linear regressions to assign proper weights to different n-grams and words within BLEU framework. These enhancements helped in improving the effectiveness of the BLEU method.

Both BLEU and NIST are widely used metrics to evaluate machine translation systems' outputs. Since they are language independent, these two methods ignore the linguistic features of the targeted natural language. A study by Qin, Wen, and Wang [13] have noticed this fact and thus used synonymous words and phrases to those found in the reference translations. In their study, a N-gram co-occurrence algorithm was used to produce pseudo translations for BLEU and NIST, the pseudo translations are based on substituting words and phrases in the reference translations for synonyms. Tests on this method have revealed clearly that the enhancement to both BLEU and NIST is more correlated to human evaluations.

In their study, Veillard, Melissa, Theodora, Racoceanu, and Bressan [14] have adopted machine learning (ML) to evaluate machine translation (MT) systems, and have proposed a new ML-based metrics, which uses support vector machine methods and include multi-class support vector machines (SVM) and support vector regression (SVR) with different kernel functions. Tests on these new ML-based metrics proved that they outperform the popular standard metrics like BLEU, METEOR, and ROUGE.

Most of the previous studies presented in this study are related to an automatic evaluation of machine translation on sentence level, where the connectivity of sentences in a document is neglected. Wong, Pun, Kit, and Webster [15] study however, is characterized by presenting a new metric to automatically evaluate the quality of the translation at a document level. They have emphasized on the structure of the outputted document by the machine translation system, more specifically, on the lexical cohesion feature. Conducted tests by the researchers showed that the adopted feature is influential and helps to improve the correlation between human judgments of machine translation outputs at the document level by 3% to 5%.

Brkic, Mikulic, and Matetic [16] have conducted a study to evaluate the machine translation (MT) from Croatian to English using two MT systems (Google Translate) and a system called LegTran that was developed and introduced by her. A reference translation conducted by a professional translator is also used. WER, PER, TER, F-measure, BLEU, and NIST as automatic evaluation methods were used. The conducted tests showed that there is no contradiction between the results of the above six methods used to identify the best MT system, except that human BLEU scores were higher than the automated BLEU score.

Condon et al. [17] study was related to the automatic evaluation of Iraqi Arabic–English speech translation dialogues. Those researchers have found that translation into Iraqi Arabic will correlate higher with human judgments when normalization (light stemming, lexical normalization, and Orthographic normalization) is used.

In their study, Adly and Al-Ansary [5] have conducted an evaluation of Arabic machine translation based on the Universal Networking Language (UNL) and the Interlingua approach for translation. The Interlingua approach relies on transforming text in the specified language into a representation form that is language independent that can be later on transferred into the target language. Three measures were used for the evaluation process; BLEU, F1 and Fmean. The evaluation was performed using the Encyclopedia of Life Support Systems (EOLSS). The effect of UNL onto translation from/into Arabic language was also studied by Alansary, Nagi, and Adly [18], and Al-Ansary [19]

The different characteristics of the Arabic language and their effect on Machine Translation were the topic of Salem, Hensman, and Nolan [6] study. In their study, the authors have proposed a model incorporating the Role and Reference Grammar technique to overcome the free word order of Arabic obstacle in the Translation process.

Carpuat, Marton, and Habash [4] study has addressed the challenges raised by the Arabic verb and subject detection and reordering in Statistical Machine Translation. To minimize ambiguities, the authors have proposed a reordering of Verb Subject (VS) construction into Subject Verb (SV) construction for alignment only which has led to an improvement in BLEU and TER scores.

A methodology for evaluating Arabic machine translation was presented in the study of Guessoum and Zantout [20]. In their study, they have evaluated lexical coverage, grammatical coverage, semantic correctness and pronoun resolution correctness. Their approach was used to evaluate four English-Arabic commercial Machine Translation systems; namely ATA, Arabtrans, Ajeeb, and Al-Nakel.

In a recent survey by Alqudsi, Omar, and Shaker [21], the issue of machine translation of Arabic into other languages was discussed. In the survey, the challenges and features of Arabic for machine translation was discussed. In addition, different approaches to machine translation and their possible application for Arabic were also mentioned in the survey. The survey concluded by indicating the difficulty of finding a suitable machine translator that could meet human requirements.

In a study by Galley, Green, Cer, Chang, and Manning [22], an Arabic-to-English statistical machine translator called the Stanford University's Arabic-to-English SMT which was built as in improvement to a previous Chinese-to-Arabic MT system was described. In their system, a comparison between three types of lexicalized reordering models was performed.

A phrase-based reordering model was used as the core engine of the system and the BLEU score was reported to have increased using their approach.

In a study by Khemakhem, Jamoussi, and Ben Hamadou [23], an Arabic-English Statistical Machine translator; called MIRCL, was discussed. The MIRCL system was built using a phrase-based approach. In addition, a solution for disambiguation of the output of the Arabic morphological analyzer was presented in their study that was used to help in selecting the proper word segments for translation purposes.

The impact of Arabic morphological segmentation on the performance of a broad-coverage English-to-Arabic Statistical machine translation was discussed in the work of Al-Haj and Lavie [24]. In their work, a phrase based statistical machine translation was addressed. Their results have showed a difference in BLEU scores between the best and worst morphological segmentation schemes where the proper choice of segmentation has a significant effect on the performance of the SMT.

III. The Methodology

This section presents the main steps, followed to accomplish this study, and summarized in Figure 1. Bilingual Evaluation Understudy (BLEU) method is adopted in this study to evaluate Babylon machine translation system and Google Translate machine translation system. The effectiveness of translation from English to Arabic using Babylon machine translation system and Google Translate system is tested using BLEU method.

In the first step we have to input 5 statements as shown below:

The source sentence in English is inputted to the machine translation system.

The translation of the source sentence using Google Translate system.

The translation of the source sentence using Babylon Translate system.

Two reference translations of the source sentence.

The second step involves the text preprocessing by dividing the text into different n-gram sizes, as follows: unigrams, bigrams, trigrams, and tetra-grams. The precision for Babylon machine translation system and Google machine translation system were computed for each of the four gram sizes. In the final step, for each of the four n-gram sizes, we compute a unified precision score for that size. These values are then compared to decide which of them get the best translation.

A. Dividing the text into different n-gram sizes

An n-gram can be defined as a sub-sequence of n items, from a given sequence of words (text or sentence). These items can be characters, words or sentences according to the application.

An n-gram can be of any number of words and each of which has a name, when the sizes of the n-grams are equal to one, two, three, or four words, they are called unigram, bigram, trigram, and tetra-gram respectively. This study deals with these types. The n-gram extraction technique to extract any size of word(s) is described in Figure 2

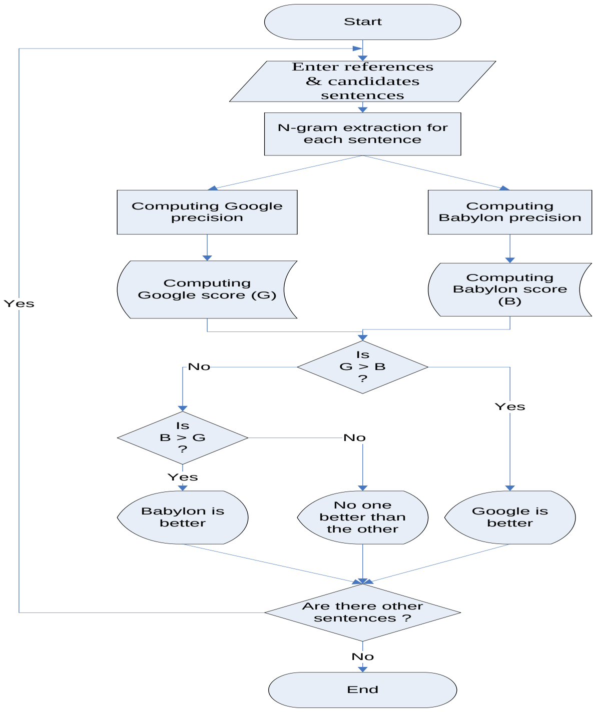

Start
Enter references & candidates sentences
N-gram extraction for each sentence
Computing Google precision
Computing Babylon precision
Computing Google score (G)
Computing Babylon score (B)
Is G > B ?
No
Yes
Is B > G ?
No
Yes
Babylon is better
No one better than the other
Google is better
Are there other sentences ?
Yes
No
End

Figure 1. Evaluation Methodology Flowchart.

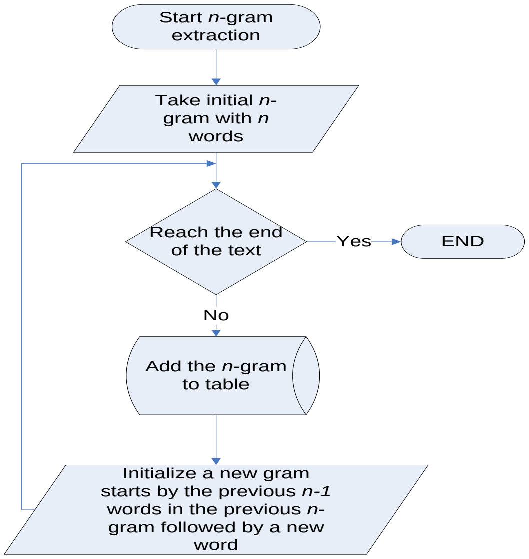

Start n-gram extraction
Take initial n-gram with n words
Reach the end of the text
Yes
END
No
Add the n-gram to table
Initialize a new gram starts by the previous n-1 words in the previous n-gram followed by a new word

Figure 2. N-grams Extraction Flowchart

To explain this method for extracting n-grams, we will provide an example for bigram size in the study; so we translate the statement “World football cup is held every four years once” into Arabic as “يـ عـ قد كـأس يـ الـ عالـ م لـ كرة الـ قدم كل اربـ ع سـ نوات مرة”, and divide it into bigrams as shown in Figure 3 below

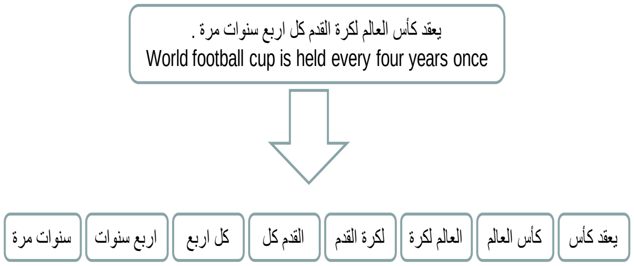

يعقد كأس العالم لكرة القدم كل اربع سنوات مرة .
World football cup is held every four years once
يعقد كأس
كأس العالم
العالم لكرة
لكرة القدم
القدم كل
كل اربع
اربع سنوات
سنوات مرة

Figure 3. Bigram Example.

B. *Babylon and Google Machine Translators Precision*

N-grams are used in many areas like information retrieval, text mining, natural language processing (NLP) … etc. In this study, n-gram extraction is used as a preprocessing technique. In order to compute the precision score for each of the four n-gram sizes, we have to count first the number of common words in every candidate and reference sentence, and then we have to divide this sum over the total number of n-grams in the candidate sentence.

To explain that, we take a source sentence as an example and translate it using Babylon machine translation system and Google Translate machine translation system, and two human translations called Reference 1 and Reference 2 as follows.

EXAMPLE 1:

Source Sentence: Banks usually lend money to persons who need it, for a specified interest.

Babylon machine translation system:

أجل من الـ مال ىإل يـ حـ تاجون الـ ذيـ ن الأ شخاص عادة تـ قدم مـ صارف
محددة مـ صالـ ح

Google Translate:

و إلـ يها، يـ حـ تاجون الـ ذيـ ن لـ لأ شخاص عادة الـ مال تـ قرض الـ بـ نوك
محددة لـ مـ صـ لحة ذلـ ك

Reference 1:

الأموال يـ حـ تاجون الـ ذيـ ن الأ شخاص بـ إقـ راض الـ بـ نوك تـ قوم
مـ عـ يـ نه فـ ائـ دة مـ قابـ ل

Reference 2:

لـ قاء لـ ها يـ حـ تاجون الـ تي الـ مـ بالـ غ الـ ناس تـ قرض عادة الـ مـ صارف
لـ لمـ صرف فـ ائـ دة

At this stage we have to compare the outputs of Google Translate system with the two references. The first comparison is based on unigram; we found that the unigrams “الـ بـ نوك Banks”, “الـ ذيـ ن who” and “يـ حـ تاجون need” are common with reference 1, also “تـ قرض lend” and “يـ حـ تاجون need” with reference 2. So, the number of common unigrams is equal to 4.

The total unigrams in output of Google Translate system for the source sentence is equal to 12. So the unigram precision is equal to (4/12) ≈ 0.33, as shown in Table 1.

Then, when we do the second comparison according to bigram, we found that “يـ حـ تاجون الـ ذيـ ن who need” bigram is the only bigram common with reference 1, with no bigrams common with reference 2. So the bigram precision is equal to (1/11) ≈ 0.09. The trigram precision and tetra-gram precision values were computed in the same way, and the results are shown in Table 1.

TABLE I. PRECISION VALUES FOR EXAMPLE 1.

MT / N-grams	Babylon machine translation system	Google Translate System
Uni-gram Precision(P1)	$\frac{3}{12}$	$\frac{4}{11}$
Bi-gram Precision (P2)	$\frac{1}{11}$	$\frac{1}{10}$
Tri-gram Precision(P3)	$\frac{0}{10}$	$\frac{0}{9}$
Tetra-gram Precision(P4)	$\frac{0}{9}$	$\frac{0}{8}$

C. *Babylon and Google Machine Translators BLEU-score*

To combine the previous precision values in a single overall score (called BLEU-score), we start by computing the Brevity Penalty (BP) by choosing the effective reference (i.e. the reference that has more common n-grams) length which is denoted by r. Then we compute the total length of the candidate translation denoted by c. Now we need to select Brevity Penalty to be a reduced exponential in (r / c) as shown in equation 1 [1]:

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{\left(1-\frac{r}{c}\right)} & \text{if } c \leq r \end{cases} \qquad (1)$$

In our example for Babylon machine translation system c = 12, r = 10, and when 12 > 10 then the BP = 1, and for Google Translate c = 11, r = 10, and when 11 > 10 then also BP = 1.

Now, we use the previous resulted BP from equation 1 to compute the final BLEU score as shown in formula (2) [1].

$$BLEU = BP \times \exp\left(\sum_{n=1}^{N} w_n \log p_n\right) \qquad (2)$$

where N = 4 and uniform weights wn = (1/N), in this study [1].

Tests on example1 showed that the BLEU score for the Babylon machine translation system is 0.075, and the BLEU score for the Google Translate is 0.115. This result indicates that Google Translate is more accurate than Babylon machine translation system, since higher BLEU score for any machine translator means that its better than its counterparts with lower BLEU scores.

Papineni, Roukos, Ward, and Zhu [1] study noted that the BLEU metric ranges from 0 to 1, where the translation that has a score of 1 is identical to a reference translation [1].

IV. THE EVALUATION

In order to speed up the calculation used in evaluating the Babylon machine translation system and the Google machine translation system we have developed a system using visual studio .Net 2008 to accomplish this goal, the main screen of the system is shown in Figure 4 as shown below.

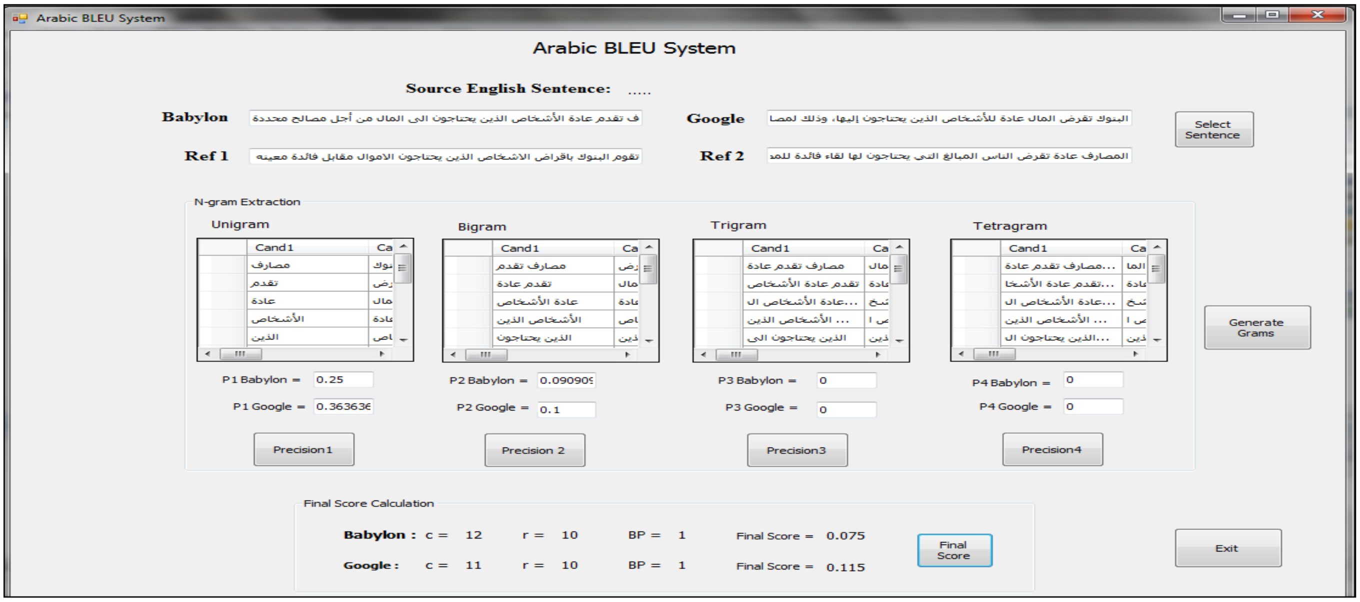
As indicated by Alqudsi, Omar, and Shaker [21], most of the approaches that have been proposed for Arabic-English machine translation was tested on limited domains; mostly news and government data. For this, to evaluate the attained results of this evaluation system, we have constructed a corpus of 100 sentences that were categorized into 7 types; past, present, future, imperative, passive, conditional “if”, and questions. In addition to that, 300 popular English sayings were also taken and translated into Arabic using both the Babylon and Google translators.

The majority of the conducted experiments on these sayings have resulted into a literal and meaningless translation of the saying. For instance, the English say “A good workman is known by his chips”; which has the Arabic meaning as “ع ند يـهان أو الـمرؤ يـ كرم الام تحان”, was literally mistranslated by both translators into “الـ شرائـ ح معروف جـ يد عامل”, as a Babylon translation, and into “لـه رقائـ ق من الـ جـ يد الـ عامل الـمعروف ومن ”, as the Google translation; which is very literal and very far from the actual meaning of the saying.

In our evaluation and testing of the translators, we have found out that in some sentences the translation precision is equal for both machine translators (Google and Babylon). However, after the application of the Arabic BLEU system on the 300 English sayings, the conducted experiments have indicated better translation accuracy by the Google translator than the Babylon translator; (0.44 for Google and 0.12 for Babylon).

It has also been noticed that Babylon translator have not succeeded in correctly translating any of the sayings at 100% accuracy, and that Google translator have succeeded; at some extent, in fully translating some of these sayings. For general translations, it has been noticed that "Google Translate system was better than Babylon machine translation system in most of the translations".

As a whole, the average precision values of Google and Babylon machine translation system for each type of sentences in the corpus are shown in Table 2 and Figure 5. It is obvious that Google Translate system was better than Babylon machine translation system.



Arabic BLEU System

Source English Sentence: .....

Babylon: ف تقدم عادة الأشخاص الذين يحتاجون الى المال من أجل مصالح محددة

Google: البنوك تقرض المال عادة للأشخاص الذين يحتاجون إليها، وذلك لمصا

Ref 1: تقوم البنوك باقراض الاشخاص الذين يحتاجون الاموال مقابل فائدة معينه

Ref 2: المصارف عادة تقرض الناس المبالغ التى يحتاجون لها لقاء فائدة للمد

N-gram Extraction

Unigram

Cand1	Ca
مصارف	نوك
تقدم	رض
عادة	مال
الأشخاص	عادة
الذين	ماص

P1 Babylon = 0.25

P1 Google = 0.363636

Bigram

Cand1	Ca
مصارف تقدم	رض
تقدم عادة	مال
عادة الأشخاص	عادة
الأشخاص الذين	ماص
الذين يحتاجون	ذين

P2 Babylon = 0.090909

P2 Google = 0.1

Trigram

Cand1	Ca
مصارف تقدم عادة	مال
تقدم عادة الأشخاص	عادة
...عادة الأشخاص ال	شخ
... الأشخاص الذين	ص ا
الذين يحتاجون الى	ذين

P3 Babylon = 0

P3 Google = 0

Tetragram

Cand1	Ca
...مصارف تقدم عادة	الما
...تقدم عادة الأشخا	عادة
...عادة الأشخاص ال	شخ
... الأشخاص الذين	ص ا
...الذين يحتاجون ال	ذين

P4 Babylon = 0

P4 Google = 0

Final Score Calculation

Babylon : c = 12 r = 10 BP = 1 Final Score = 0.075

Google : c = 11 r = 10 BP = 1 Final Score = 0.115

Figure 4. The Main Screen of the Arabic BLEU System.

TABLE II. AVERAGE PRECISION FOR EACH TYPE OF SENTENCES TYPE

Type / Translator	Past	Present	Future	Imperative	Passive	Conditional "if"	Questions
Babylon machine translation system	0.193	0.206	0.172	0.196	0.239	0.146	0.205
Google machine translation system	0.386	0.414	0.267	0.404	0.273	0.163	0.294

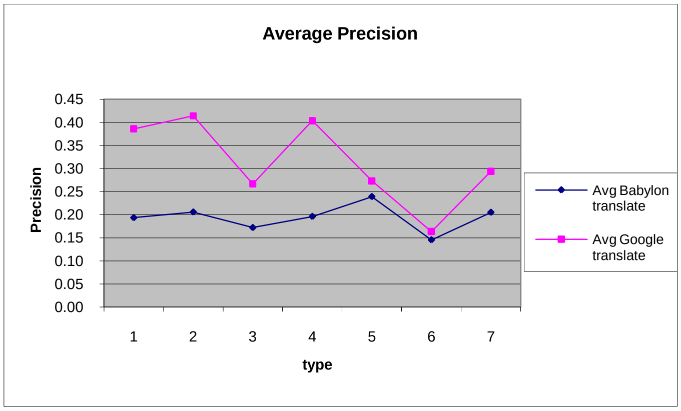


Figure 5: Summary of Average Precision

We have noticed from the conducted experiments that the translation quality of 21% of translated English sentences into Arabic using Babylon translate system were more accurate than Arabic sentences outputted by Google Translate system. While the translation quality of Google Translate system was better than the translation quality of Babylon translate system to translate 69% English sentences into Arabic. The two machine translators yield equal accuracy to translate 10% of the English sentences into Arabic.

V. CONCLUSION

English-to-Arabic machine translation has been a challenging research issue for many of the researchers in the field of Arabic Natural Language Processing. In this study, we have evaluated the effectiveness of two automatic machine translators that could be used for English-to-Arabic translation and vice versa. The used machine translators are Google machine translator and the Babylon machine translator.

The accuracy of any machine translator is usually evaluated by comparing the results to human judgments. There is no standard Arabic corpus that can be used for such evaluations, for this we had to collect our data from different Internet websites representing two types of data; a set of well-known English sayings and a set of sentences that were translated manually by two human translators for judgment purposes.

Although the collected data was of small size, the well-known English sayings usually presented a challenge for the Machine translators into Arabic.

After applying our developed Arabic BLEU System on the collected data, we have found out that the overall translation precision for Google was 0.314 and the overall translation precision for the Babylon machine translation system was 0.194. As for the English popular sayings, it has been found out that the Google translate system has better accuracy than that of Babylon translation system (0.44 for Google and 0.12 for Babylon). Based on these findings, we can conclude that the Google Translate system is better than Babylon machine translation system for the translation from English into Arabic.

Furthermore, we have found out that Babylon machine translation system was incapable of translating some of the English words into Arabic properly. For example, the Babylon machine translator could not fully translate the following English sentence: "Great talkers are little doers ", since the outputted Arabic translation was: "يراقب top talkers كبيرة
المحدثين قليلة ".

VI. Future Work

Measures of translation quality based on exact matching of word forms are of challenge because of the orthographic variation; which is especially severe in the Arabic language. To solve such problem, and as a future research, we are planning to find a technique to solve it. Other automatic evaluation methods for machine translators like NIST, METEOR, ROUGE and RED will be included in our future studies.

We have tested our experiments on a small size of data, as part of the future work we are planning on collecting more data and perform tests using the new data as well as any available standard data that could be found.

References

[1] Papineni, K., Roukos, S., Ward, T. and Zhu, W.J. 2002. "BLEU: a method for automatic evaluation of machine translation". In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02). Stroudsburg, PA, USA, pp. 311-318.

[2] Doddington G. 2002. "Automatic evaluation of machine translation quality using n-gram co-occurrence statistics". In Proceedings of the second international conference on Human Language Technology Research (HLT '02). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 138-145.

[3] Al Dam, R.; Guessoum, A. 2010. "Building a neural network-based English-to-Arabic transfer module from an unrestricted domain," In Proceedings of IEEE International Conference on Machine and Web Intelligence (ICMWI), pp.94-101.

[4] Carpuat M., Marton Y., and Habash N., 2010. "Improving Arabic-to-English Statistical Machine Translation by Reordering Post-verbal Subjects for Alignment", In Proceedings of the ACL 2010 Conference Short Papers, pp. 178–183, Uppsala, Sweden.

[5] Adly, N. and Alansary, S. 2009. "Evaluation of Arabic Machine Translation System based on the Universal Networking Language", In Proceedings of the 14th International Conference on Applications of Natural Language to Information Systems “NLDB 2009”, pp. 243-257.

[6] Salem Y., Hensman A., and Nolan B. 2008, "Towards Arabic to English Machine Translation", ITB Journal, Issue 17, pp. 20-31.

[7] Riesa, J., Mohit, B., Knight, K., Marcu, D. 2006. "Building an English-Iraqi Arabic Machine Translation System for Spoken Utterances with Limited Resources", In the Proceedings of INTERSPEECH, Pittsburgh, USA.

[8] Nießen S., Och F.J., Leusch G., Ney H.. 2000. "An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research". In Proceedings of the 2nd International Conference on Language Resources and Evaluation, pp. 39-45.

[9] Melamed D., Green R., and Turian J.P., 2003. "Precision and recall of machine translation". In Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: companion volume of the Proceedings of HLT-NAACL 2003--short papers - Volume 2 (NAACL-Short '03), pp. 61-63.

[10] Palmer, D.D. 2005. "User-centered evaluation for machine translation of spoken language,", Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.5, pp. v/1013- v/1016.

[11] Akiba, Y.; Imamura, K.; Sumita, E.; Nakaiwa, H.; Yamamoto, S.; Okuno, H.G.; 2006, "Using multiple edit distances to automatically grade outputs from Machine translation systems," Audio, Speech, and Language Processing, IEEE Transactions on , vol.14, no.2, pp. 393- 402.

[12] Yang M., Zhu J., Li J., Wang L., Qi H., Li S., Daxin L. 2008. "Extending BLEU Evaluation Method with Linguistic Weight,", 2008. ICYCS 2008. The 9th International Conference for Young Computer Scientists, pp.1683-1688.

[13] Qin Y., Wen Q., Wang J., 2009. "Automatic evaluation of translation quality using expanded N-gram co-occurrence,", NLP-KE 2009. International Conference on Natural Language Processing and Knowledge Engineering, pp.1-5.

[14] Veillard, A.; Melissa, E.; Theodora, C.; Racoceanu, D.; Bressan, S. 2010. "Support Vector Methods for Sentence Level Machine Translation Evaluation,", 22nd IEEE International Conference on Tools with Artificial Intelligence (ICTAI), vol.2, pp.347-348.

[15] Wong, B.T.M., Pun, C.F.K., Kit, C., Webster, J.J. 2011. "Lexical cohesion for evaluation of machine translation at document level," 7th International Conference on Natural Language Processing and Knowledge Engineering (NLP-KE), 2011, pp.238-242.

[16] Brkic, M. Mikulic, B.B. ; Matetic, M. ; Basic Mikulic, Bozena; Matetic, Maja; 2012. "Can we beat Google Translate?,", Proceedings of the ITI 2012 34th International Conference on Information Technology Interfaces (ITI) , pp.381-386.

[17] Condon S., Arehart M., Parvaz D., Sanders G., Doran C. and Aberdeen J. 2012. "Evaluation of 2-way Iraqi Arabic–English speech translation systems using automated metrics", Machine Translation, Volume 26, Nos. 1-2, pp. 159-176.

[18] Alansary S., Nagi M. and Adly N. 2009. "The Universal Networking Language in Action in English-Arabic Machine Translation", In Proceedings of 9th Egyptian Society of Language Engineering Conference on Language Engineering, (ESOLEC 2009), Cairo, Egypt.

[19] Alansary, S. 2011. "Interlingua-based Machine Translation Systems: UNL versus Other Interlinguas", in Proceedings of the 11th International Conference on Language Engineering, Cairo, Egypt.

[20] Guessoum A. and Zantout R. 2005. "A Methodology for Evaluating Arabic Machine Translation Systems", Machine Translation, issue 18, pp. 299-335.

[21] Alqudsi A, Omar N., and Shaker K. 2012, "Arabic Machine Translation: a Survey", Artificial Intelligence Review (July 2012), pp.1-24

[22] Galley M., Green S., Cer D., Chang P.C., and Manning C.D. 2010, " Stanford University’s Arabic-to-English Statistical Machine Translation System for the2009 NIST MT Open Evaluation", in NIST Open Machine Translation Evaluation Meeting.

[23] Khemakhem I., Jamoussi S., Ben Hamadou A., 2010, "The MIRACL Arabic-English Statistical Machine Translation System for IWSLT 2010", in Proceedings of the 7th International Workshop on Spoken Language Translation, Paris, France.

[24] Al-Haj H. and Lavie A., 2012, " The Impact of Arabic Morphological Segmentation on Broad-coverage English-to-Arabic Statistical Machine Translation ", vol. 26, no. 1-2, pp. 3-24.

AUTHORS PROFILE

Mohammed Al-Kabi Mohammed Al-Kabi, born in Baghdad/Iraq in 1959. He obtained his Ph.D. degree in Mathematics from the University of Lodz/Poland (2001), his masters degree in Computer Science from the University of Baghdad/Iraq (1989), and his bachelor degree in statistics from the University of Baghdad/Iraq(1981). Mohammed Naji AL-Kabi is an assistant Professor in the Faculty of Sciences and IT, at Zarqa University. Prior to joining Zarqa University, he worked many years at Yarmouk University in Jordan, Nahrain University and Mustanserya University in Iraq. He also worked as a lecturer in Jordan University of Science and Technology. AL-Kabi's research interests include Information Retrieval, Web search engines, Data Mining, Software Engineering & Natural Language Processing. He is the author of several publications on these topics. His teaching interests focus on information retrieval, Web programming, data mining, DBMS (ORACLE & MS Access).

Taghreed M. Hailat, born in Irbid/jordan in 1986. She obtained her MSc. degree in Computer Science from Yarmouk University (2012), and her bachelor degree in Computer Science from Yarmouk University (2008). Currently, she is working at Irbid chamber of Commerce as a Computer Administrator and previously as a trainer of many computer courses at Irbid Training Center.

Emad M. Al-Shawakfa is an Assistant Professor at the Computer Information Systems Department at Yarmouk University since September 2000. He was born in Jordan in 1964 and holds a PhD degree in Computer Science from Illinois Institute of Technology (IIT) – Chicago, USA in the year 2000, a MSc in Computer Engineering from Middle East Technical University in Ankara-Turkey in the year 1989, and a BSc in Computer Science from Yarmouk University in Irbid-Jordan in the year 1986. His research interests are in Computer Networks, Data Mining, Information Retrieval, and Arabic Natural Language Processing. He has several publications in these fields and currently working on others

Izzat M. Alsmadi. is an associate professor in the CIS department at Yarmouk University, Irbid, Jordan. Born in Jordan 1972, Izzat Alsmadi had his master and PhD in software engineering from North Dakota State University (NDSU), Fargo , USA in the years 2006 and 2008 respectively. His main areas of research include: software engineering, testing, metrics, and information retrieval**.**

Parametric and Non Parametric Time-Frequency Analysis of Biomedical Signals

S. Elouaham, R. Latif, A. Dliou
ESSI, National School of applied Sciences, Ibn Zohr University
Agadir, Morocco

M. LAABOUBI
High School of technology
Ibn Zohr University
Guelmim, Morocco

F. M. R. Maoulainie
Team of Child, Health and Development, CHU, Faculty of Medicine, Cadi Ayyad University,
Marrakech, Morocco

***Abstract*— Due to non-stationary multicomponent nature of the electrocardiogram (ECG) signal, its analysis by the monodimensional techniques, temporal and frequenctial, can be very difficult. The use of the time-frequency techniques can be inevitable to achieve to a correct diagnosis. Between the different existing parametric and non-parametric time-frequency techniques, the Periodgram, Capon, Choi-Williams and Smoothed Pseudo Wigner-Ville were chosen to deal with analysis of this biomedical signal. In a first time, a comparison between these time-frequency techniques was made by analyzing modulated signal to make in evidence the technique that gives a good resolution and low level of cross-terms. In a second time, the Periodogram which presents a powerful technique was applied to a normal and abnormal ECG signal. The results show the effectiveness of this time-frequency in analyzing this type of biology signal.**

***Keywords-** ECG; Time-frequency; Periodogram; Capon; Choi-williams.*

I. INTRODUCTION

The electrocardiogram (ECG) is one of the major physiological signals generated from heart's rhythmic polarization and depolarization. The P and T waves and QRS complex are the principles components of this biological signal. The P wave presents the arterial activation, the QRS complex reflects the ventricular depolarization and T wave refers to the ventricular repolarization [1, 2]. These different ECG components have specific shapes in the time and frequency domains, which allow the specialists to detect possible existing anomalies only by analyzing ECG signal in these two domains. The electrocardiograms signals are generally non-stationary. The information giving by using the Fourier transform does not provide any indication about the frequency notion over time. Otherwise, the temporal domain presents the weakness that the frequency component which presents the pertinent information is not visualized. To exceed the one-dimensional limitations, the use of the time-frequency techniques is inevitable; these techniques provide an accurate description of the non-stationary behavior of the ECG signals, as a conjoint representation, in time and frequency. The time-frequency techniques used are able to present the evolution of ECG signal power as a function of both time and frequency. In this study, the two major categories of the time-frequency techniques are utilized, the parametric and the non-parametric ones. The chosen time-frequency techniques are selected for their interesting proprieties [1-15]. The used time-frequency techniques are: Periodogram, Capon, Choi-williams and Smoothed pseudo Wigner-Ville. A comparison based on the variance and the resolution factors of some modulated signals are proposed to select the most suitable technique, which will be more adequate for analyzing the ECG signal. In this paper, the work was based for analyzing a normal and an abnormal ECG with some types of arrhythmia pathology. Cardiac arrhythmias are divided into two groups, the first one is life threatening and requires immediate treatment with an automated external defibrillator (AED) or an implantable cardioverter defibrillator (ICD) [16, 17]. The second one is not deadly, but requires treatment sustainability. The abnormal ECG signals used in this work are obtained from patients with cu ventricular tachyarrhythmia and malignant ventricular arrhythmia.

The Periodogram performance surpass others techniques in order to reducing the cross terms and increasing the resolution. The cross-terms present the principle problem of the time-frequency which can obscure some important information in the time-frequency plan.

The article is organized as the follows, firstly we will present, in the section II, the different techniques used in this work. Secondly, in the section III, we will detail the biomedical signals used in this study. Thirdly, the section IV and V will show the main obtained results. Finally, we will conclude with a conclusion in the section VI.

II. TIME-FREQUENCY TECHNIQUES

A. Parametric techniques

The parametric time-frequency techniques used in this work are the Capon, the Periodogram.

1. Capon distribution

The estimator of minimum variance called Capon estimator (CA) does not impose a model on the signal. At each frequency f, this method seeks a matched filter whose response is 1 for the frequency f and 0 everywhere else [3-7].

$$\begin{aligned} CA(n,f) &= a(n,f)^{H} R_{X} a(n,f) \\ &= \frac{1}{Z_{f}^{H}.R_{X}[n]^{-1}.Z_{f}} \end{aligned} \qquad (1)$$

Where:

- $CA(n,f)$ is the output power of the filter Capon, excited by the discrete signal x(n) sampled at the period te,

- $a(n,f)=(a_0,...,a_p)$ is the impulse response of the filter at frequency n,

- $R_x[n]=E\{x[n]x^T[n]\}$ is the autocorrelation matrix of crossed x(n) of dimension $(p+1)*(p+1)$,

- $x[n]=(x(n-p),...,x(n))$ is the signal at time n,

- $Z_f^H=(1,e^{2i\pi fte},...,e^{2i\pi ftep})$ is the steering vector,

- (p+1) is the number of filter coefficient and the exponent H for conjugate transpose and the superscript T for transpose.

Periodogram technique

The Periodogram (PE) is the derivate of the Capon (CA) technique. The spectral estimator of this method is defined by the following equation [3-7]:

$$PE(n,f)=Z_f^H.R_x.Z_f/((\mathrm{p}+1)^2) \qquad (2)$$

The two previous techniques defined by the equations 1 and 2 can be applied sliding windows. There is no theoretical criterion for choosing the filter order and duration of the window [4]. The parametric techniques depend on the signal so that the frequency response has a different shape and then different properties according to the signal characteristics. The choice of the window is more crucial to the time-frequency resolution. CA and PE estimator usually has a better frequency resolution. Both techniques are well suited to signals containing some strong spectral components such as ECG biomedical signals.

B. *Non parametric techniques*

1. Smoothed Pseudo Wigner-Ville

The majority of the non parametric time-frequency representations are represented by the cohen class [7-15]. This class includes inter alias the Wigner-Ville distributions. Among the large number of existing time-frequency representations some authors have proposed using the Wigner-Ville [9, 10]. To avoid the covering of frequential components in the time-frequency representation, in this study we use in the place of the real signal $x(t)$ the analytical signal $x_a(t)$ defined by the expression:

$$x_a(t)=x(t)+iH\{x(t)\} \qquad (3)$$

Where i^2 =-1 and $x(t)$ is the signal with real values and $H\{x(t)\}$ it's Hilbert transform.

The distribution of Wigner-Ville associated to a signal $x(t)$, of finished energy, is the function $W_x(t,f)$ depending of the temporal (t) and frequential (f) parameters. This distribution is given by the following expression:

$$W_x(t,f)=\int_{-\infty}^{+\infty}x\left(t+\frac{\tau}{2}\right).x^*\left(t-\frac{\tau}{2}\right)e^{-2i\pi f\tau}d\tau \qquad (4)$$

Where x^* indicates the complex conjugate of $x(t)$.

The transform called the Smoothed Pseudo Wigner-Ville (SPWV) is implanted in this work to attenuate the interference terms presented between the inner components figured in Wigner-Ville representation which decrease the visibility of the time-frequency image. The SPWV use two smoothing windows h(t) and g(t). These smoothing windows are introduced into the Wigner-Ville distribution definition in order to allow a separate control of interference either in time (g) or in frequency (h). The expression of this representation is defined by [12]

$$SPWV_x(t,f)=\int_{-\infty}^{+\infty}h\left|\left(\frac{\tau}{2}\right)^2\right|\int_{-\infty}^{+\infty}g(t-u)x\left(u+\frac{\tau}{2}\right) \qquad (5)$$
$$.x^*\left(u-\frac{\tau}{2}\right)e^{-2i\pi f\tau}dud\tau$$

Where h(t) is a smoothing frequential window and g(t) is a smoothing temporal window.

2. Choi-Williams technique

The Choi-Williams distribution was developed as a method of finding the Wigner-Ville Distribution with the minimum amount of cross-term interference. The Choi-Williams distribution is a shift-invariant transform. Essentially, the distribution is a smoothed version of the Wigner-Ville distribution through a kernel function defined by [15- 17]:

$$f(\varepsilon,\tau)=\exp\left[-\frac{(\pi\varepsilon\tau)^2}{2\sigma^2}\right] \qquad (6)$$

The Choi-Williams distribution is then defined as:

$$CW_x(t,f)=\sqrt{\frac{2}{\pi}}\int\int_{-\infty}^{+\infty}\frac{\sigma}{|\tau|}\exp^{-2\sigma^2(s-t)^2/\tau^2}A_x\exp^{-j2\pi f\tau}dsd\tau \qquad (7)$$

Where

$$A_x=x\left(s+\frac{\tau}{2}\right)x^*\left(s-\frac{\tau}{2}\right) \qquad (8)$$

The smoothing of the distribution is controlled by the constant σ. As $\sigma\to\infty$, the Choi-Williams (CW) distribution will simplify converges to the Wigner-Ville distribution, as the kernel goes to 1. The cross-terms are reduced with smaller values of σ. The Choi-Williams distribution is a bilinear time-frequency representation [17].

III. Biomedical Signals

A. *Normal ECG signal*

The normal signal ECG is characterized with P wave, T-wave, and complex the QRS. The figure 1 shows an example of a normal ECG signal.

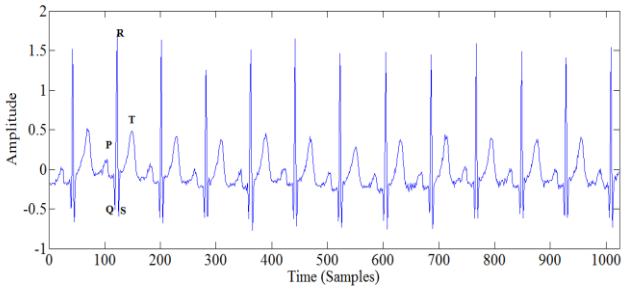


Figure.1. Normal ECG signal

B. *Abnormal signal*

The abnormal signals used in this work are cu ventricular tachyarrhythmia, and malignant ventricular arrhythmia. The figure 2 represents misalignments of the third QRS complex of cu ventricular tachyarrhythmia.

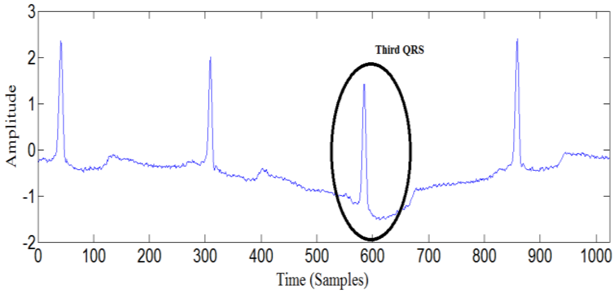


Figure. 2. Abnormal ECG signal with cu ventricular tachyarrhythmia

And the last anomaly is malignant ventricular arrhythmia presented in figure 3. The T-wave signal has disappeared and the QRS complexes are wide.

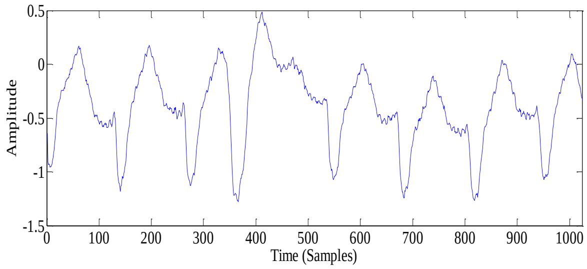


Figure. 3. Abnormal ECG signal of malignant ventricular

The ECG signals were provided by MIT-BIH database [18].

IV. Performance Of Time-Frequency Techniques

A. *Modulated signals*

The time-frequency techniques are applied in this section to the signal given by the following equation,

$$x(t) = \sin(2\pi\ (f_1 + 50t)t) + \sin(2\pi\ (f_2 + 50t)t) + \sin(2\pi\ t)\sin(2\pi\ (f_3 + 50t)t) \quad (9)$$

Where

f_1=50, f_2=100, f_3=110 ;

The signal used is composed of three components; two components of high frequency and the third component is of low frequency .The goal of this application is the choice of appropriate method for specific application

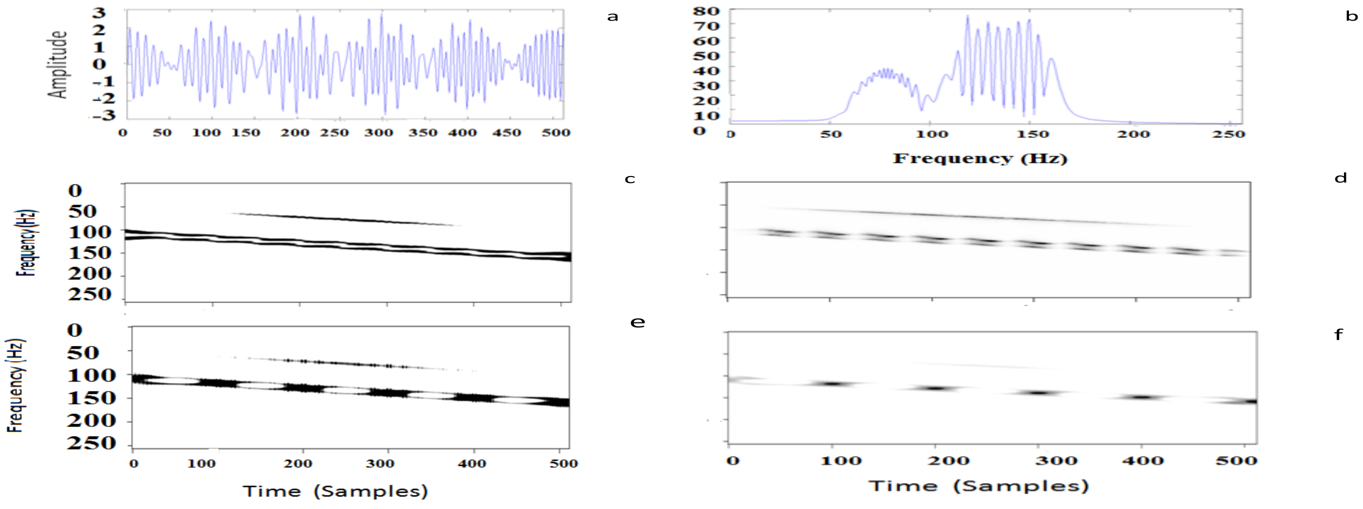


Figure. 4. Modulated signal (a), Fourier spectrum (b), PE (c), CA (d), CW (e), SPWV (f) Resultants

The figure 4a presents the modulated signal used for the comparison between the time-frequency techniques. The Fourier transform spectrum presented in figure 4b shows the evolution of frequency content from 50 to 165, and cannot give any information about number of the components frequency and their change over time. The figure 4c presented by periodogram gives a good localization of the two high frequency components than the others techniques. This technique can separate clearly the frequency components. The non parametric time-frequency images, figures 4e and 4f, show the presence of high level of cross-terms and low degree of resolution which make the identification of the two frequency components more delicate than the periodogram technique.

B. *Variance*

The time-frequency techniques used in this study are applied to a monocomponent signal to find the most performance technique.

The monocomponent signal used is given by the following equation:

$$x(t) = ae^{j\phi(t)} \tag{10}$$

The instantaneous frequency (IF) is given by the following equation:

$$f = \frac{1}{2\pi} d\phi / dt = f_0 + \beta t \tag{11}$$

Where a=1, f_o=0.05f_s, β = 0.4f$_s$, $\phi(t)$ is the analytic signal phase and f_s = 1/T is the sampling frequency.

The bias (B) and the variance (VAR) of the estimate present the most important factors that decide the quality of estimation. These two notions can be defined by the following expressions:

$$B\left(\hat{f}_i(t)\right) = \varepsilon\left[\Delta \hat{f}_i(t)\right] \tag{12}$$

$$VAR\left(\hat{f}_i(t)\right) = \varepsilon\left[(\Delta \hat{f}_i(t))\right]^2 \tag{13}$$

With:

$$\Delta \hat{f}_i(t) = f_i(t) - \hat{f}_i(t)$$

Where $f_i(t)$ and $\hat{f}_i(t)$ are the instantaneous frequency and instantaneous frequency estimate respectively. The signal length used in the time-frequency techniques is N=256 samples and the total signal duration is 1 s. The sampling frequency was f_s=2 NHz. Using different Signal-to-Noise Ratio (SNR), gaussian white noise samples are added to the signal. The figure 5 shows the performance of the PE, CA and SPWV, CW time-frequency techniques applied to a linear FM signal with 256 points.

The table 1 gives the variance values of the parametric (PE, CA) and non-parametric (SPWV, CW) techniques. According to the results of this table, the Periodogram time-frequency technique has a minimal variance for all SNR's. The low minimum variance can indicate the performance of the time-frequency techniques. The PE technique surpasses the other time-frequency techniques in robustness where it gives the minimum variance at low SNR.

V. Results And Discussion

The figures 6 to 8 present the time-frequency images of the normal and abnormal ECG signals (figure 1 to 3). This time-frequency images are obtained by using the calculation of the equation 2 and 10 of the Periodogram and the Choi-Williams techniques respectively. The figures 6a (2D) and 6a’ (3D) show the Periodogram images of the normal ECG signal (figure 1). By the same way, the figures 6b (2D) and 6b’ (3D) expose the Choi-willams images of the same signal.

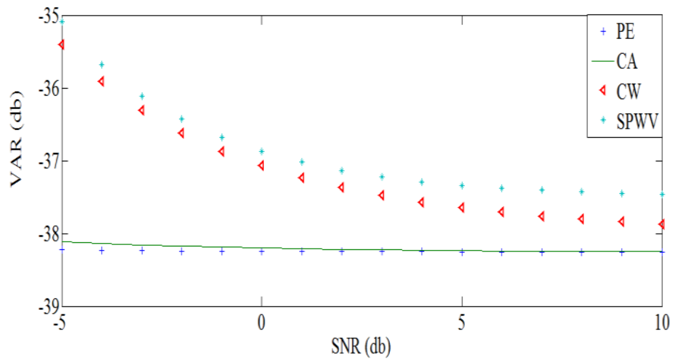


Figure.5. Performance of PE, CA, CW and SPWV in IF estimation of a linear FM signal with length N = 256 samples.

TABLE: VARIANCE OF PARAMETRIC AND NON PARAMETRIC TECHNIQUES

SNR(db)	PE	CA	CW	SPWV
-5	-38.2252	-38.1106	-35.4000	-35.0960
-4	-38.2318	-38.1379	-35.9079	-35.6796
-3	-38.2370	-38.1598	-36.3039	-36.1092
-2	-38.2411	-38.1776	-36.6174	-36.4273
-1	-38.2444	-38.1920	-36.8683	-36.6749
0	-38.2471	-38.2038	-37.0703	-36.8680
1	-38.2491	-38.2136	-37.2340	-37.0191
2	-38.2507	-38.2216	-37.3674	-37.1354
3	-38.2520	-38.2282	-37.4772	-37.2246
4	-38.2530	-38.2336	-37.5682	-37.2911
5	-38.2538	-38.2381	-37.6440	-37.3417
6	-38.2544	-38.2419	-37.7074	-37.3795
7	-38.2548	-38.2451	-37.7606	-37.4081
8	-38.2552	-38.2477	-37.8052	-37.4300
9	-38.2555	-38.2500	-37.8429	-37.4473
10	-38.2556	-38.2519	-37.8749	-37.4615

The Periodogram time-frequency images (figure 6a et 6a’) give a good location of the QRS complexes and the T waves. By the comparison of the periodogram results with those obtained by Choi-williams (Fig. 6b and 6b'), we note that the periodogram technique gives the best results. The figure 7 presents the time-frequency images of the abnormal ECG signal obtained from a patient with cu ventricular tachyarrhythmia (figure 2). The Periodogram time-frequency image (fig. 7a) gives us a good localization of QRS complexes with the transient’s appearance in the third QRS complex. The visualization of these transients in the time-frequency image in the figure 7a’ indicates the ability of detecting abnormalities by the Periodogram technique. The Periodogram technique can track the change in the frequency components of each QRS complex. The figures 7b and 7b’ represent the Choi-williams time-frequency images. These images give no information about the transients in the third QRS complex which present the existing abnormality. These obtained results expose the robustness of the Periodogram technique for identifying the pathology than the Choi-williams method.

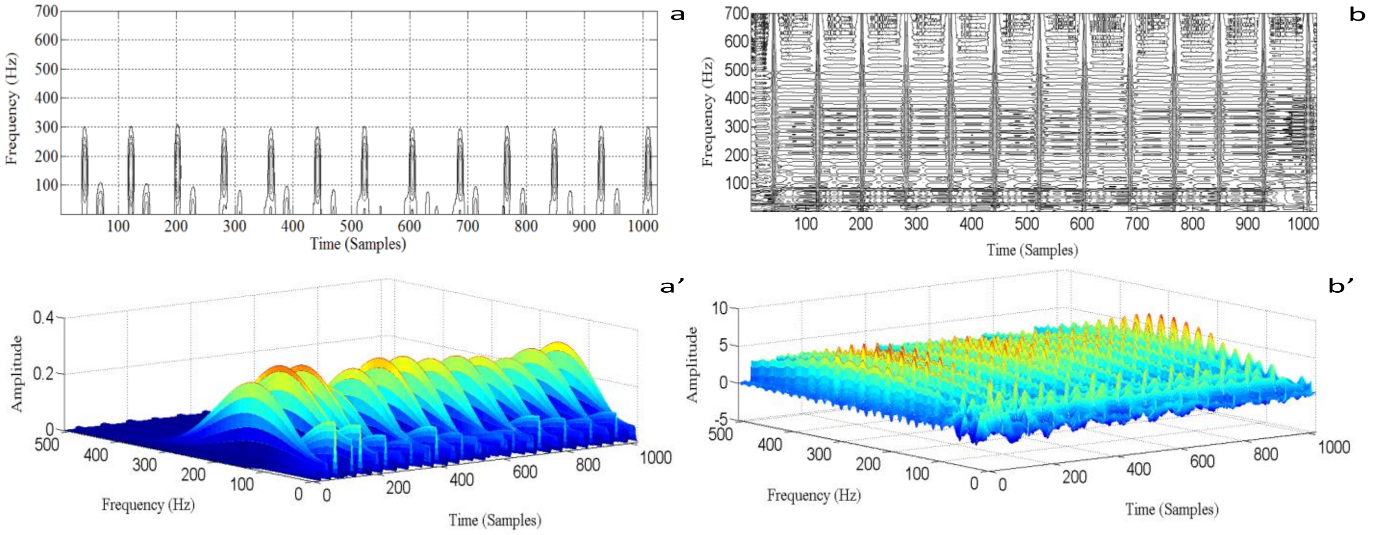


Figure.6. Periodogram (a and a') and Choi-williams (b and b') time-frequency images of a normal ECG signal.

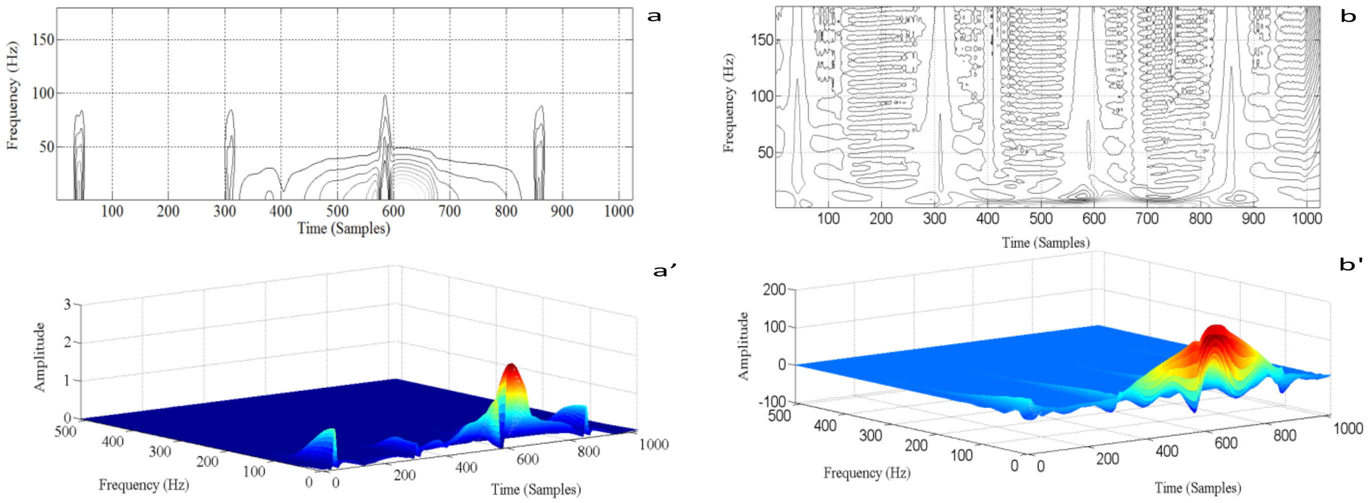


Figure.7. Periodogram (a and a') and Choi-williams (b and b') time-frequency images of an abnormal ECG signal with cu ventricular tachyarrhythmia

The figure 8 presents the time-frequency images of the abnormal ECG signal obtained from a patient with malignant ventricular arrhythmia (figure 3). The QRS complexes in this signal are wide and abnormal, while the T wave has disappeared. The signal seems to have an irregularity and is changing over time. The Periodogram parametric time-frequency images show a good localization of QRS complex and absence of T waves. All frequency components in the signal of Fig. 3 are clearly revealed by the Periodogram technique (fig 8a and 8a'). However, The figures 8b and 8b' obtained from the application of the Choi-williams non-parametric technique present the interference terms which hide some useful information. The Choi-williams technique can't track the changes of the frequency components the QRS complex of the abnormal signal in the case the malignant ventricular arrhythmia. The high performance of the PE technique is clearly demonstrated in the figure 8 (a and a'). We note from the results of the figure 8 that the parametric Periodogram technique are better than the non parametric Choi-williams technique for analyzing this delicate ECG abnormality.

The Periodogram parametric technique allow us to identify many types of abnormal ECG signals and reveal changes of frequency from the detection of the T waves and the QRS complexes. This parametric technique seems to be a good approach for analyzing normal and abnormal ECG signals compared to the non-parametric time-frequency techniques.

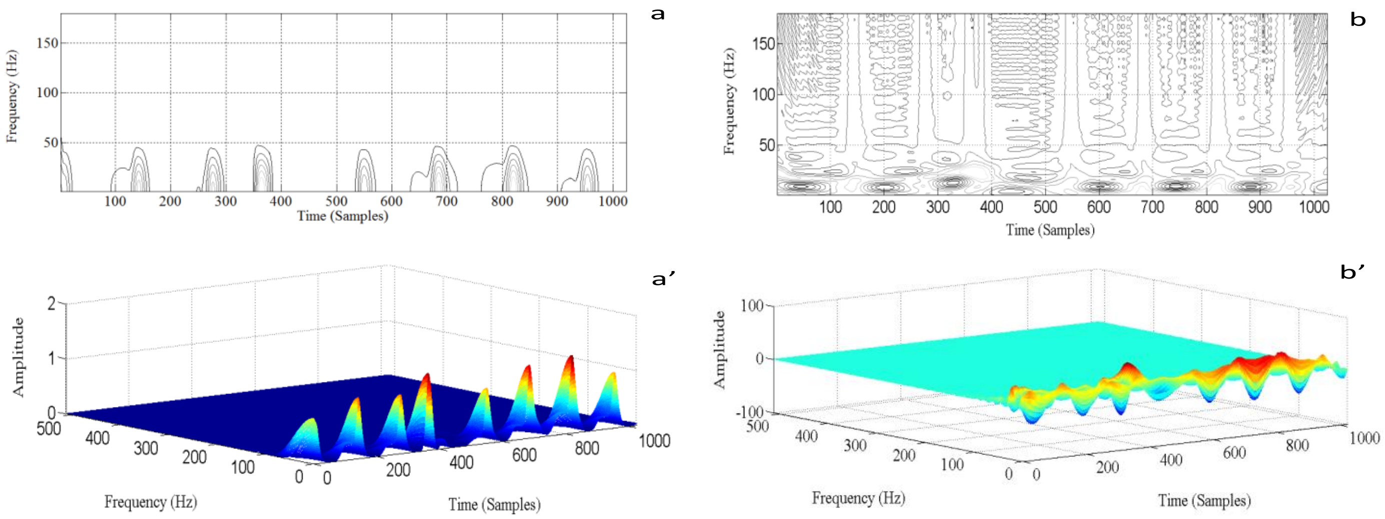


Figure.8 : Periodogram (a and a') and Choi-williams , (b and b') time-frequency images of an abnormal ECG signal with malignant ventricular arrhythmia

VI. Conclusion

This paper presents a comparative study of some parametric and non parametric techniques. These time-frequency techniques are applied to normal and abnormal biomedical ECG signals. The abnormalities treated in this work, are cu ventricular tachyarrhythmia and malignant ventricular arrhythmia.

For the cu ventricular tachyarrhythmia case, the parametric technique Periodogram allows us to detect clearly the transients and QRS complexes with good resolution in the time-frequency plan. For malignant ventricular arrhythmia case, the obtained results expose the ability and the effectiveness of the parametric technique Periodogram to reveal the non-stationary behavior of this type of abnormal signals. The parametric technique can track changes in the frequency components of each QRS complex in time. The results obtained in this work, show the performance of the parametric technique Periodogram for analyzing biological signals such as ECG signal compared to others parametric and non-parametric time-frequency techniques.

References

[1] P. de Chazal, Maraia O' Dwyer, Richard B. Reilly, "Automatic classification of heartbeats using ECG morphology and heartbeat interval features", IEEE Transactions on Biomedical Engineering 51 (7) pp. 1196–1206, 2004.

[2] G Bortolan, Christian Brohet, Sergio Fusaro, "Possibilities of using neural networks for ECG classification", Journal of Electrocardiology 29 Suppl:10-16, 1996.

[3] M.T. Özgen, "Extension of the Capon's spectral estimator to time–frequency analysis and to the analysis of polynomial-phase signals". Signal Process. 83 (3), pp. 575–592, 2003

[4] F. Castanié, Spectral Analysis Parametric and Non-Parametric Digital Methods（Ltd, 2006）.

[5] S. Elouaham, R. Latif, A. Dliou, E. Aassif, B. Nassiri " Analyse et comparaison d'un signal ECG normal et bruité par la technique temps-frequence parametrique de capon ," Conference Mediterranéene sur l'Ingenierie Sure des Systemes Complexes [MISC'11, ENSA Agadir, Maroc, Mai 2010].

[6] S. Elouaham, R. Latif, A. Dliou, F. M. R. Maoulainine, M. Laaboubi, " analysis of biomedical signals by the empirical mode decomposition and parametric time-frequency techniques " [International Symposium on security and safety of Complex Systems, may 2012, Agadir, Morocco].

[7] P. Goncalves F. Auger, P. Flandrin. Time-frequency toolbox, 1995.

[8] P. Flandrin, N. Martin, M. Basseville, Methodes temps- frequence. Suppl. Trait. Signal 9, 1992, pp. 77-147.

[9] Labib M. Khadra, "The smoothed pseudo-Wigner ville distribution in speech processing ", Int. J. Electronics , Vol. 65, No 4, pp. 743-755, 1988.

[10] R. Latif, E. Aassif, G. Maze, A. Moudden, B.Faiz, "Determination of the group and phase velocities from time-frequency representation of Wigner-Ville", Journal of Non Destructive Testing & Evaluation International, Vol. 32, 7, pp. 415-422, 1999.

[11] R. Latif, E. Aassif, G. Maze, D.Decultot, A. Moudden, B. Faiz, " Analysis of the circumferential acoustic waves backscattered by a tube using the time-frequency representation of wigner-ville ", Journal of Measurement Science and Technology, Vol. 11, 1, pp. 83-88, 2000.

[12] R. Latif, E. Aassif, A. Moudden, B. Faiz, "High resolution time-frequency analysis of an acoustic signal backscattered by a cylindrical shell using a Modified Wigner-Ville representation", Meas. Sci. Technol.14, pp. 1063-1067, 2003.

[13] R. Latif, E. Aassif, A. Moudden, B. Faiz, G. Maze , 'The experimental signal of a mullayer structure analysis by the time-frequency and spectral methods," NDT&E International, Vol. 39, Issue 5, pp. 349-355, 2006.

[14] R.Latif , M. Laaboubi , E. Aassif , G. Maze. "Détermination de l'épaisseur d'un tube élastique à partir de l'analyse temps-fréquence de Wigner-Ville", Journal Acta-Acustica, Vol 95, Number 5, pp. 843-848, 2009.

[15] A. Dliou, R.Latif, M. Laaboubi, F. M. R. Maoulainine, "Arrhythmia ECG Signal Analysis using Non Parametric Time-Frequency Techniques" International Journal of Computer Applications (0975 – 8887) Vol 41– No.4, 2012.

[16] A. Dliou, R.Latif, M. Laaboubi, M. Laaboubi, F. M. R. Maoulainine, "Abnormal ECG Signal Analysis using Non Parametric Time-Frequency Techniques" Arabien Journal of Science engineering, July 2012 (Accepted).

[17] H. Choi, and W. Williams. "Improved time-frequency representation of multicomponent signals using exponential kernels", IEEE Trans. on Acoustics, Speech and Signal Processing, vol. 37, pp. 862-871. 1989.

[18] Physiobank, Physionet, Physiologic signal archives for biomedical research, http://www.physionet.org/physiobank.

Validating Utility of TEIM: A Comparative Analysis

Rajesh Kulkarni [1]
Department of Computer Engineering
Bhiavarabai Sawant College of Engineering and Research
Pune, India

P.Padmanabham [2]
Department of Computer Science and Engineering
Bharat Institute of Engineering and Technology
Hyderabad, India

***Abstract*— Concrete efforts to integrate Software Engineering and Human Computer Interaction exist in the form of models by many researchers. An unconventional model called TEIM (The Evolved Integrated Model) of Software Engineering and Human Computer Interaction was proposed by us. There is a need to establish correlation with prior models as well validate utility of TEIM. In this paper product PS designed using SE-HCI integration model TEIM is evaluated by making a comparative analysis. For evaluation UGAM and IOI tools designed by DR.Anirudha Joshi are used. Our analysis showed that correlation of TEIM exists with prior models. Regression analysis showed that high correlation exists between TEIM and prior model.**

Keywords- SE; HCI; UGAM; IOI; PS; TEIM.

I. Introduction

Better user experience is an oft expressed quality of the products designed nowadays. Many efforts in this regard lead to various proposals of smooth integration of SE(software engineering) processes with HCI(human computer integration) for product development were done [1], [3], [4], [5], [8], [10], [11], [12]. We got inspired by these and designed a product application by name PS(Personal Secretary) using SE-HCI integration model of [1] and adding empathy map [7], [9] to it. The steps used for designing PS evolved into a new SE-HCI integration model by name TEIM [2].

II. Validating Utility Of TEIM

Dr.Anirudha Joshi in [1] had proposed UGAM (Usability Goals Achievement Metric) to measure user experience goals and IOI (Index of Integration) to measure extent of integration of HCI activities in SE processes.

We used UGAM and IOI to evaluate PS in this paper. Section III explains UGAM score calculation for PS, section IV explains IOI score calculation for PS and section V explains mathematical and comparative analysis of PS vis-à-vis TEIM.

For mathematical and comparative analysis statistical methods of regression, Pearson’s coefficient, ANOVA are used.

III. Usability Goals Achievement Metric

Usability Goals Achievement Metric (UGAM) proposed by [1] is a product metric that measures the quality of user experience.

A. UGAM components [1]

- Goals: High level user experience goals.
- Goal parameters: Goals divided in to goal parameters.
- Weight: Weights are in the range 0-5 indicating least relevant to most relevant.
- Score: Scores are in the range 0 to 100 broken down to four categories 0- worst user experience, 25- bad, 50- undecided state, 75- good and 100- best.

UGAM calculation for TEIM Model is in Table 1.UGAM parameter labels are in Figure 1.The average weight assigned is 2.8 which is in the range 2.4 to 3.4.As per UGT (Usability Goal Setting Tool) the weight assigned is balanced.UGAM tool proposed by Joshi. A. et al., [1] is used to measure user experience of PS designed by us.

PS was designed using TEIM model [2] (refer Figure 5). TEIM [2] evolved as an unconventional model of integrating software development process with usability aspects [1][3] wherein we were trying to understand SE-HCI integration efforts of Joshi .A [1] , Ferre[3][4], Seffah[5], designed PS using their techniques of integration and adding our beliefs.

PS was evaluated on teaching staff of Computer Engineering dept. of BSCOER, Pune by us and scores were assigned. UGAM was calculated [1] using the formula $\sum (W_p \text{ X } S_p)/ \sum W_p$ where Wp is the weight of the goal parameter p and S_p is the score of the goal parameter p.

IV. UGAM And IOI Relationship

In [1] data from industry projects was available in the form of 61 industry projects UGAM and IOI scores .We could not get access to such data so our reference data were the UGAM and IOI scores of Joshi. A. et al. [1].

Using this reference data and extended waterfall model[8] we used the same techniques [1] of evaluation for establishing relationship between UGAM and IOI as well relationship between our UGAM + IOI scores vs. [1] scores .Methods used to establish correlation between and their results are as followed:

- Pearson’s Correlation: Refer Table III for Pearson Coefficient calculation and A for the results.
- Linear Regression: Refer Table IV, V for Linear Regression calculation and B, C for results.
- ANOVA: Refer Table VI, VII and D for ANOVA calculations and results respectively.

A. Pearson's Interpretation

Interpretation of Pearson's Correlation results: a positive Coefficient indicates values of variable A vary in the same direction as variable B. Characterizations of Pearson r:

- .9 to 1 very high correlation
- .7 to .9 high correlation
- .5 to .7 moderate correlation
- .3 to .5 low correlation

Very high positive correlation exists between the Variation of UGAM and the variation of IOI.There is a significant positive correlation (r= 0.99, $p < 0.0005$ two-tailed) between UGAM and IOI $r_{xy} = 1$, adjusted $r_{xy} = 0.99$.

All the above techniques including the plot drawn for UGAM vs. IOI (refer Figure 3, 4 and F) validate linear correlation between UGAM and IOI. Also Table VIII, F and E establish a linear correlation between TEIM and [1].

A: learn ability: A1: find ability: easy to find option, A2: take less time to learn, A3: able to learn on their own, A4: product: internally consistent, A5: consistent with other products, A6: consistent with earlier version, A7: retain critical/infrequent tasks
B: speed of user: B1: ability to do tasks easily all times, B2: ability to navigate quickly/easily, B3: not load user user's memory, B4: flexibility: control seq of tasks, B5: complete tasks in less effort, B6: automatic personalization, B7: localized for specific market, B8: user ability to customize,
C: Ease of Use, C1: interface communicate model, C2: predict next step intuitively, C3: No entry barrier: complete tasks, C4: No unnecessary tasks, C5: automate routine tasks, C6: product: always on/accessible
D: Ease of Communication, D1: Information Architecture well categorized, D2: clear understanding of text/visuals,
E: Error-free use, E1: should give good feedback/status, E2: Should not induce errors, E3: Errors: tolerate/forgive/prevent, E4: Help to recover from errors
F: Subjective Satisfaction, F1: Feel in control/behavioral appeal, F2: Emotional engagement/fun/appeal, F3: Aesthetical/Visceral appeal, F4: Average weight

Figure 1. UGAM PARAMETER LABELS

B. Regression Coefficient

$R^2 = ((1/N)*\sum [(X1-\bar{x}) *(Y1-\bar{Y})]/(\sigma_x * \sigma_y))^2 =1$

σ_x= sqrt $[\sum(X1-\bar{x})^2/N]$=11.83

σ_y = sqrt$[\sum (Y1-\bar{Y})^2/N]$= 8.72

Adjusted $R^2 = 0.99$

IOI significantly determines the scores of UGAM with predictor IOI accounting for 99% of the variance in UGAM (adjusted $R^2 = 0.99$)

C. Linear Regression

TABLE I. UGAM CALCULATION FOR TEIM MODEL

	goals and goal parameters	weights	goal para meter score	goal score	UGAM score
A	A1	3	75	52.5	
	A2	4	75		
	A3	3	50		
	A4	3	75		
	A5	3	50		
	A6	0	0		
	A7	4	0		
B	B1	2	75	50	43.15
	B2	2	50		
	B3	3	75		
	B4	2	50		
	B5	2	75		
	B6	3	0		
	B7	3	75		
	B8	2	0		
C	C1	3	25	25	
	C2	2	75		
	C3	3	0		
	C4	3	75		
	C5	2	0		
	C6	5	0		
D	D1	3	25	25	
	D2	4	25		
E	E1	3	50	40	
	E2	3	50		
	E3	2	25		
	E4	2	25		
F	F1	3	50	60	
	F2	3	50		
	F3	4	75		
	F4	2.8			

Regression equation form $\bar{Y} = b0 + b1*x$

$$b1 = \sum ((X1- \bar{x})*(Y1-\bar{Y}))/\sum(X1- \bar{x})$$
$$b0 = \bar{Y} - b1* \bar{x}$$
$$\bar{Y} = b0 + b1*x$$

$$\bar{Y} = 14.95 + 0.74 * x$$

D. Anova Results

According to F Sig/Probability table with df(2,1) F must be at least 19.000 to reach $p< 0.05$. So F score is statistically significant. Hence our hypothesis is supported.

E. RK VS AJ Correlation

The range of correlation coefficient is -1 to 1. Since our result is 0.99 or 99%, it means the variables have a high positive correlation.

TABLE II. IOI CALCULATION FOR TEIM MODEL

Phases	HCI activities	Recommended weights	weights	Activity Score	Phase Score	IOI Score
A	A1	3 to 4	3	75	54.55	46.74
	A2	2	4	50		
	A3	1 to 3	3	50		
	A4	1 to 3	1	25		
B	B1	4 to 5	4	50	25	
	B2	4 to 5	4	25		
C. Construction	C1	3	3	75	68.75	

A. Communication: A1.Contextual User Studies/modeling, A2: Ideation with multidisciplinary team, A3.Product definition/Information Architecture/Wireframes, A4. Usability evaluation, refinement
B. Modeling: B1.Detailed UI prototyping, B2.Usability Evaluation, refinement
C. Construction: C1: Development Support reviews by Usability team, C2: Usability Evaluation (Summative)

Figure 2. IOI parameter labels

TABLE III. PEARSON'S COEFFICEINT

	X	Y			
group	ugamscore	ioiscore	XY	X^2	Y^2
rk	43.15	46.74	2016.83	1861.92	2184.63
aj	66.81	64.17	4287.208	4463.58	4117.79
SUM	109.96	110.91	6304.02	6325.50	6302.42
n	2				

TABLE IV. LINEAR REGRESSION

		X1	Y1
srno	entity	ugam	ioi
1	RK	43.15	46.74
2	AJ	66.81	64.17
	SUM	109.96	110.91
	MEAN	54.98	55.455

F. UGAM Vs IOI Calculation

The closer the points come to straight line stronger the relationship. We will express the strength of the relationship between 0 and 1.

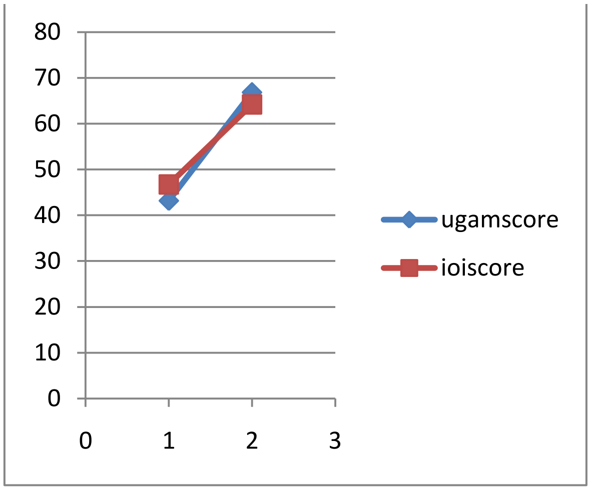


Figure 3. UGAM VS IOI CORRELATION

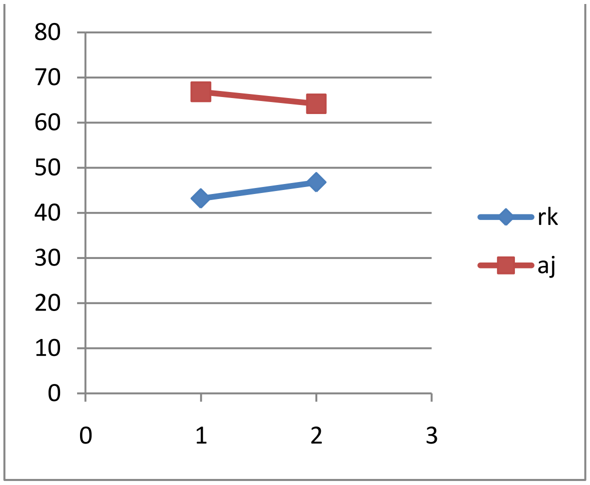


Figure 4. VS AJ RK

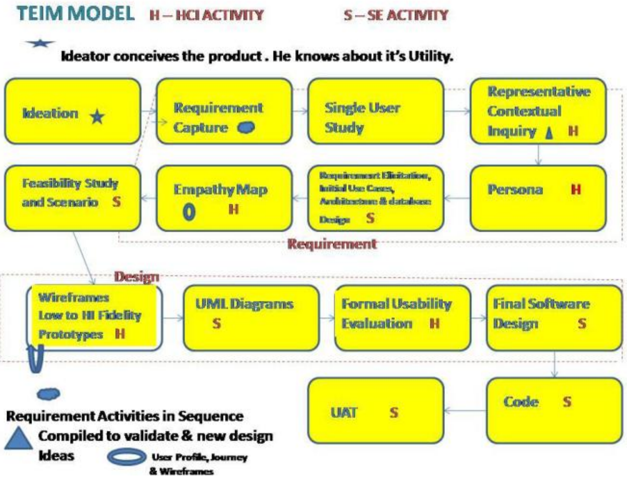

TEIM MODEL
H – HCI ACTIVITY
S – SE ACTIVITY
Ideator conceives the product . He knows about it's Utility.
Ideation
Requirement Capture
Single User Study
Representative Contextual Inquiry H
Feasibility Study and Scenario S
Empathy Map H
Requirement Elicitation, Initial Use Cases, Architecture & database Design S
Persona H
Requirement
Design
Wireframes Low to HI Fidelity Prototypes H
UML Diagrams S
Formal Usability Evaluation H
Final Software Design S
UAT S
Code S
Requirement Activities in Sequence
Compiled to validate & new design Ideas
User Profile, Journey & Wireframes

Figure 5. TEIM MODEL

TABLE V. LINEAR REGRESSION

srno	X1-x̄	Y1-Ȳ	$(X1-\bar{x})^2$	$(Y1-\bar{Y})^2$		
	A	B	C	D	A*B	C*D
1	-11.83	-8.72	139.95	75.95	103.10	10629.29
2	11.83	8.72	139.95	75.95	103.10	10629.29
sum			279.90	151.90	206.20	

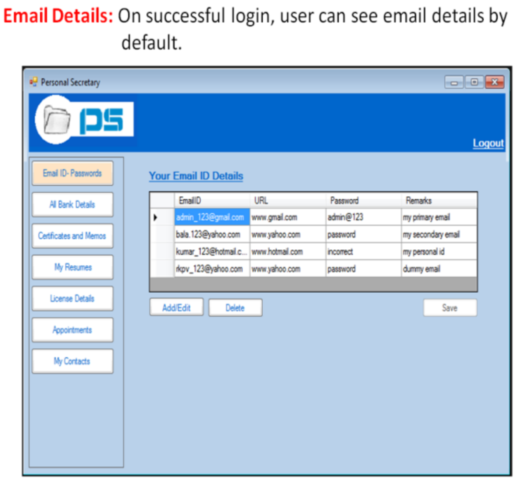
TABLE VI. ANOVA CALCULATION

SOURCE	SS	DF	MS	F
AMONG	422.10	2	211.05	21.26
WITHIN	9.93	1	9.93	

SSTOTAL	432.03
R^2	0.98

TABLE VII. ANOVA CALCULATION

	X_1	X_2	$(X_1)^2$	$(X_2)^2$
	43.15	66.81	1861.92	4463.58
	46.74	64.17	2184.63	4117.79
∑	89.89	130.98	4046.55	8581.37
$(\sum x)^2$	8080.212	17155.76		
M	44.945	65.49		



Email Details: On successful login, user can see email details by default.

Your Email ID Details

EmailID	URL	Password	Remarks
admin_123@gmail.com	www.gmail.com	admin@123	my primary email
bala.123@yahoo.com	www.yahoo.com	password	my secondary email
kumar_123@hotmail.c...	www.hotmail.com	incorrect	my personal id
rkpv_123@yahoo.com	www.yahoo.com	password	dummy email

Figure 6. PS Product Screen

V. CONCLUSION

We designed product PS (refer Figure 6) getting inspired from prior work of integration of Human Computer Interaction and Software Engineering processes also adding our own beliefs such as empathy map [7], [9]. Whatever design steps we applied we compiled them together as a new integration model of SE and HCI and called it as TEIM- The Evolved Integration Model of SE and HCI [2]. Dr. Anirudha Joshi's work in this area is here [8]. Dr. Anirudha Joshi's tools UGAM and IOI were used to calculate UGAM score (43.15) and IOI score (46.74) respectively for the product PS. Though scores were on lower side as compared to [1] (beta version of PS was tested) they showed linearity and strong correlation.

ACKNOWLEDGMENT

We thank Dr. Anirudha Joshi who through his HCI monsoon course and research motivates many of us to take plunge in SE-HCI research. We thank teaching and non-teaching staff of BSCOER for helping us in designing and evaluating PS.

REFERENCES

[1] Joshi. A.,et al., "Measuring Effectiveness Of HCI Integration in software developmentprocesses"J.syst.Software(2010),doi:10.1016/j.jss.2010.03.078

[2] R.Kulkarni, P. Padmanabham, "TEIM- The Evolved Integrated Model Of SE And HCI" UNIASCIT, Vol 2 (3), 2012, 301-304

[3] X. Ferre, N. Juristo, H. Windl, and L. Constantine, "Usability Basics for Software Developers", IEEE Software, vol.18, no. 1, pp. 22-29, January/February 2001.

[4] Xavier Ferré, Natalia Juristo Juzgado, Ana María Moreno, ―Improving Software Engineering Practice with HCI Aspects. SERA 2003: 349 363.

[5] Seffah A. et al, ―Human-Centered Software Engineering ― Integrating Usability in the Software Development Lifecycle‖, Human – Computer Interaction Series, 2005, Volume 8, I, 3-14, DOI: 10.1007/1-4020-4113-6_1

[6] http://www.idc.iitb.ac.in/~anirudha/pdfs/IoI%20v3_0.pdf.

[7] E.Chen, "Agile User Experience Design Techniques to get you from an Idea to a prototype in 180 minutes", Workshop UXUTSAV, Bangalore, India, 16-17 June 2012.

[8] A. Joshi,"Integration of Human-Computer Interaction Activities in Software Engineering for Usability Goals Achievement", Thesis Submitted for the Degree of Doctor of Philosophy, INDIAN INSTITUTE OF TECHNOLOGY BOMBAY, 2011.

[9] http://www.xplane.com

[10] Xavier Ferré, Natalia Juristo Juzgado, Ana María Moreno,"Improving Software Engineering Practice with HCI Aspects" SERA 2003: 349 363

[11] Jerome, B., Kazman, R. "Two Solitudes.‖ Human Centered Software Engineering." Ed. Ahmed Seffah, Jan Gulliksen and Michel C. Desmarais. Netherlands: Springer, 2005. 59-69.

[12] R. Kazman, J. Gunaratne, B. Jerome,"Why Can't Software Engineers and HCI Practitioners Work Together?". Human-Computer Interaction Theory and Practice - Part 1 (Proceedings of HCI International ‗03), (Crete, Greece), June 2003, 504-508. Nielsen, Jakob (1992): The Usability Engineering Life Cycle. In IEEE Computer, 25 (3) pp. 12-22.

AUTHORS PROFILE

Rajesh Kulkarni, He received BE, Mtech degrees in 1995 and 2005, respectively. He wrote 01 book and published 3 journal papers.

P.Padmanabham, He is a double Post-Graduate in Engineering & Technology (*M.Tech-Computer Science and M.Tech-Advanced Electronics*) and Ph.D in Computer Science & Engineering.He has Over 40 years of experience in Technical Education and in the areas of Teaching, Administration, Research . He was Team Leader for Impact Evaluation for MHRD/WORLD BANK "TEQIP-1" project. He wrote 04 books and 10 papers.

Collaborative System Model for Dynamic Planning of Supply Chain

Latifa Ouzizi [1], EL Moukhtar Zemmouri [2],
Youssef Aoura [3], Hussain Ben-azza [1]
[1] Dept. of Industrial Engineering
[2] Dept. of Mathematics and Computer Sciences
[3] Dept. of Material and Process
Laboratoire de Modélisation Mathématique et Informatique
ENSAM-Meknès, University My Ismail
B.P. 15290 EL Mansour, Meknes 50500, Morocco

***Abstract*— The business need to be structured as an integrated supply chain pushes companies to make use of a greater level of co-operation and coordination. As a means of coordination, negotiation has been chosen in this work. The object of this paper is to present formalism for negotiation in dynamic planning of a supply chain with the objective of maximizing the overall profit of each partner. To model the SC, we use the multi agent approach. Each enterprise is represented by its negotiator agent. The negotiations are formalized using UML language. The proposed negotiation process allows agents to develop a feasible production schedule.**

Keywords- supply chain; negotiation; collaboration; dynamic planning; UML

I. Introduction

In the collaborative planning supply chain context, there is a need firstly to develop sophisticated optimization and decision support tools to help explore and analyze alternatives and secondly to develop tools for coordination and collaboration.

How can supply chain partners form temporary alignments to quickly respond to market requirements as well as effectively utilize their competencies? The ability of partners to plan quickly and effectively, utilize their resources throughout the chain is a key to successful supply chain planning. To achieve this, it is crucial to rapidly and effectively coordinate them through the planning process, where various constraints must be taken into account such as capacity, quality, cost, timeliness, and inter-dependencies between partners. The constraint may be limited to a partner (e.g., the capacity of production, panes machines…), or related to more than one, (e.g., the quantity of the components to be purchased, a manufacturing service should be scheduled to start after the procurement service is completed). A solution to one partner (called a node of supply chain) does not have a global view and would not satisfy both intra-node and inter-nodes constraints.

A solution of a node is usually unable to take into account the constraints embedded in interdependencies among the partners, very often resulting in incoherent and contradictory hypotheses and actions. Existing studies on this problem have focused on facilitating bilateral exchange between customers and suppliers, and have relied on complete information about resources and tasks without adequately capturing the dynamics and uncertainties of the operating environments [1]. It is a complex problem to schedule, and coordinates a set of partners from a large number of resources under various constraints and even uncertainties. The complexity is mainly due to the ambiguity in determining the requirements of components node’s; the uncertainty of solutions to component services (e.g., availability, capacity, and cost); and interdependencies among component services. The uncertainties and constraints may result in dynamisms and difficulties in searching and coordinating the services. Given this observation, the main problem is to find a way to achieve coordination and coherence among the decisions of partners in a supply chain network.

Agent technology helps understand and model complex real-world problems and systems by concentrating on high-level abstractions of autonomous entities [1,2]. The benefits of adopting agent technology in supply chains have been recognized in an increasingly wide variety of applications involving inter-enterprise collaboration, extending the boundaries of strategic partnership to wherever the network technologies can reach.

The application of multi-agent systems (MAS) in manufacturing and supply-chain management is not new. In intelligent manufacturing, agents have been used in the following functional areas: manufacturing control [3], collaborative design [4] and coordination in MAS for agile manufacturing [5]. Montreuil et al. [6] have developed a strategic network for supply chain. Lin and Solberg [7] have developed a market mechanism to coordinate agents in real time in an integrated shop floor control model. Sikora and Shaw [8] have provided a multi-agent framework for the coordination and integration of information systems.

Cauvin et al. [9] proposed an approach to minimize the impact of disruptive events on the whole intra-organizational information system. It is based on an analysis of disruptive events. The aim of this work is to assist human decision makers in the design of the recovery process, proposing them solutions for the final decision. These functionalities are characteristics of a decision corrective system. The system is unable to perform autonomous corrective actions.

Lorena et al [10] complete [9]'s study. Their system performs autonomous corrective control actions but not propose how agents make decisions.

Compared to previous approaches, we will model supply chain as a multi agent system and each partner as a multi agent system able to communicate and negotiate and plan in uncertain environment.

In our study, we suppose that each partner of the supply chain has established the planning for the future horizon. Our aim is to propose a framework for information sharing and negotiation for a collaborative planning when there were unexpected orders or impervious change.

II. FRAMEWORK'S SUPPLY CHAIN USING MULTI AGENTS SYSTEM

A. *Conceptual model of the SC*

As a solution to the problem of collaboration for planning supply chain, partners use agent negotiation. Due to dynamic changes in the internal and external environments, it is not easy to coordinate the conflicts of interests among supply chain members. What's more, quick response to those dynamic changes is required. Coordination of activities across a network of suppliers is essential for reacting quickly to uncertain environments [11]. For this reason, the use of an agent system has come to the fore. An agent system uses a coordination mechanism to approach a global optimization, along with the local objective of each agent. In addition, negotiations are widely being used as a coordination mechanism [12].

We model a supply chain as a number of nodes. A node can be supplier or\and customer of other nodes. Each node is a multi-agent system composed of a planner agent and a negotiator one. The communication and negotiation is established between negotiator agents via messages. We distinguish message of orders, agreements, proposition and negotiation.

We present on figure 1, the conceptual model of the proposed architecture using UML language.

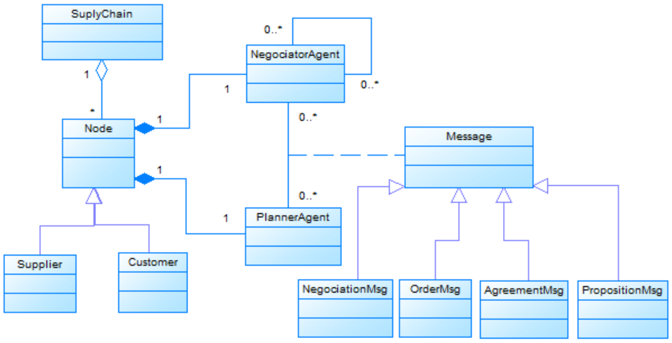


Figure 1. Conceptual model of a supply chain

B. *Architecture of the node of the supply chain*

As showed in figure1, a node of supply chain is composed of a planner agent (PA) and a negotiator agent (NA). The planner agent has tools for planning and make decision while the negotiator agent have to first, communicate and negotiate with customers and suppliers; second to communicate with its planner agent.

As shown in Figure 2, this approach allows separating planning tasks and communication protocol. The NA negotiates with suppliers and customers and communicates with the PA. The description of negotiation process will be presented in section III.

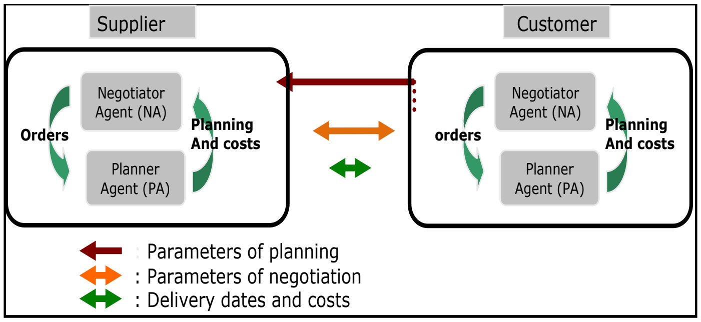


Figure 2. Architecture of a node of supply chain (NSC) in relation with customer or supplier

We assume that the planning of each partner of the SC is established. In case of an unexpected order, internal event or changing the supply plans, The planner agent has to calculate new planning and gives new solutions of planning. The PA can work in two modes:

- Backward mode: when there is an unexpected order from customers or an internal change of production system. In this case, the PA calculates the components it needs to ask from its suppliers, negotiates with them and gives new scheduling. Each schedule gives the production quantities produced by periods and the overall cost of the solution in terms of over hours and the quantities contracted.
- Direct mode: when one or more suppliers give new supply plan different from what it is provided. In this case, the PA must calculate the amount of product that can realize from these components and negotiate with customers.

We describe in next sections tools that assist the PA for planning and decision.

C. *Tools for planning and decision of the planner agent*

1) *Linear programming model of the planner agent*

A planner agent encapsulates the knowledge needed to perform every time planning and costs study. When it receives a modification of constraints curves from customers or suppliers, it uses a specific planning technique depending on the problem at hand. An example of a linear model used by the planner agent is presented.

The planning model consists in minimizing the various production costs (in terms of normal working hours, or overtime hours), the unemployed hours and costs of subcontracting and penalties of delay.

In addition, it is assumed that:

- Procurements made for period t-1 will be useful for the production of the period t
- Products take exactly one unit of time to be moved between two nodes (i.e. what is produced, stored or moved by a node for period t can be used by the following node for the period t+1, without worrying about the exact moment when transformations are carried out for the period t)
- Times of change of manufacture (set-up times) are negligible
- The production is carried out in a complex workshop in which the production lines are in series (in the event of parallel production lines, the capacity of the resources are multiplied)
- The process plans are known as well as the occupancy rate of each product on each machine. In our case, our interest is on purchased items at the entry, times spent on each machine and finally products at the exit, without worrying about intermediate products

First, let us denote the notations used. Then, the planning model will be presented.

2) Notations used

a) Definition of sets:

P: Set of products indexed by p (1<=p<=Np)

Npi: Number of products that are supplied (raw materials) indexed by i (1<=i<=Npi)

Npo: Number of products that are delivered indexed by j (Npi+1 <= j <= Np)

R: Set of resources indexed by r

T: The planning horizon

T': The planning horizon plus one period

Various costs taken into account:

$cpn_{j,r}$: Production cost in normal hour of one unit of product j on the resource r

$cot_{r,t}$: Cost of one overtime hour on the machine r during the period t

$csc_{j,t}$: Cost to subcontract one unit of product j during the period t

$chu_{r,t}$: the cost of one unemployed hour on the machine r during the period t

Information concerning the production system:

$g_{i,j}$: Quantity of product i required to manufacture one unit of product j

$C_{r,t}$: Normal capacity of production, in hours, of the resource r for the period t

$b_{j,r}$: Time to produce one unit of the product j on the resource r

MaxSCj,t: Maximum sub-contracted quantity of product j during the period t

MaxOHr,t: Maximum of overtime hours on the resource r during the period t

t0: First period of the plan

b) Information concerning customers and suppliers:

$CGDS_{i,t}$: quantity of product i proposed by suppliers at the beginning of period t (This quantity is a known parameter that corresponds to a strict constraint when computing a direct planning or a soft constraint when computing a backward planning.

$CGUS_{j,t}$: quantity of product j asked by customers at the end of period t (This quantity is a known parameter that corresponds to a strict constraint when computing a backward planning or a soft constraint when computing a direct planning.)

$ct_{j,c,t}$: The unit penalty (in €/day) of the product j not delivered in time (differed delivery) to the customer c at the period t

Secondary variables:

$CDS_{i,t}$: quantity of item i supplied at period t (is equal to $CGDS_{i,t}$ for direct planning)

$CUS_{j,t}$: quantity of product j delivered to customers at period t (is equal to CGUSj,t for backward planning)

c) Decision variables:

$X_{j,t}$: Quantity of products j produced during the period t in normal hours.

$SC_{j,t}$: Quantity of products j subcontracted for the period t.

$XHS_{j,t}$: Quantity of products j produced in overtime at the period t.

$U_{r,t}$: Unemployed hours on the resource r at the period t.

$Late_{j,t}$: Delay of delivery of the product j at the period t.

D. Objective function

To take into account the effect of time, i.e. the reliability of the data that the node has, an up-dating rate (1/(1+α)t-t0) is introduced, where α is an actualization factor. The function is:

$$Min \left[\begin{array}{l} \sum_{t=t0}^{T} \left(\frac{1}{(1+\alpha)^{t-t0}} \left[\sum_{j \in Npo} \sum_{r \in R} X_{j,t} * cpn_{j,r} + (\mathrm{csc}_{j,t} * SC_{j,t}) + ct_{j,c,t} * late_{j,t} \right.\right. \\ \left.\left. \quad + \sum_{r \in R} (\cot_{r,t} * b_{j,r} * XHS_{j,t} + chu_{r,t} * U_{r,t}) \right] \right. \end{array} \right] \quad (1)$$

E. Constraints

The planning model constraints are for each i ∈Npi, j ∈Npo, t ∈ T and r ∈R.

$$CDS_{i,t} \leq CGDS_{i,t} \qquad (2)$$

$$(X_{j,t} + XHS_{j,t} + SC_{j,t}) = CUS_{j,t} \qquad (3)$$

$$\sum_{j \in Npo} (X_{j,t} + XHS_{j,t})^{*} g_{i,j} = CDS_{i,t} \qquad (4)$$

$$late_{j,t} \geq (CGUS_{j,t} - CUS_{j,t}) \qquad (5)$$

$$late_{j,t} \geq 0 \qquad (6)$$

$$\sum_{j \in Npo} b_{j,r}{}^{*}\, XHS_{j,t} \leq MaxO_{r,t} \qquad (7)$$

$$SC_{j,t} \leq MaxSC_{j,t} \qquad (8)$$

$$\sum_{j \in Npo} b_{j,r}{}^{*}\, (X_{j,t} - XHSj,t) + U_{r,t} = C_{r,t} \qquad (9)$$

$$CDS_{i,t}, CUS_{j,t}, XHS_{j,t}, X_{j,t}, U_{r,t}, SC_{j,t} \geq 0 \qquad (10)$$

Explanations of constraints are:

2: The provisioning necessary for the production at period t should not exceed what is proposed by the suppliers.

3: At each period t and for each product j, the quantity delivered is equal to the sum of what is produced in normal hours, overtime or subcontracting.

4: At each period t, the quantity of component i used in production is equal to the sum of the requirement in components at period t.

5 and 6: The delay of the delivery of product j at the period t is equal to the maximum of 0 and the difference between what is required and what is produced.

7: Limits overtime available to a maximum value.

8: Limit subcontracted quantities for each product with a maximum value.

9: Ensures that the capacity of the resources available is equal to the sum of the operational durations and the unemployed hours minus overtime.

10: Indicates that all the variables of decision are positive or null.

III. THE MODEL OF NEGOTIATION PROCESS IN THE SC

A. Negotiation in supply chain

The negotiation is the communication process of a group of agents to reach a mutual agreement accepted by all parties [13]. For example, in the field of production, the agreement could focus on quality, costs and deadlines. Therefore, the basic idea behind the negotiation is to reach a consensus. Negotiation can be competitive or co-operative behavior by different agents involved in it [14].

The strategy of negotiation that we propose is based on a co-operative negotiation. The flow of information shared between the different agents of the supply chain includes initial orders, agreements and propositions.

We distinguish the nodes of the SC in relation with customers and the internal nodes of the chain. We will not study the case of nodes in relation with external suppliers because it is assumed that there are no problems with raw material suppliers because in general they work on stock.

B. The process of planning and negotiation in the supply chain

In the case of a change due to unexpected orders, the negotiator agent (NA) seeks urgent solutions. Firstly, it checks the inventory, if the order can be fulfilled, it responds with an agreement else it asks the planner agent to restart schedule. In the first case, the AP uses backward plan, calculate components needs, which the NA asks to suppliers (figure 4) and gives solutions that can satisfy the customer with additional costs ie additional costs of raw materials, costs of over times, quantities that can be outsourced and possible delays. It transmits the message to the (NA) which in turn transmits it to the customer. If there is agreement, the negotiation ends else the customer can change requests. The message can be represented by a 6-upla (product, quantity, time, price, penalties, possible delays on certain products). The NA keeps the database of different scenarios proposed by the PA for a final decision.

In the case of changes in supply plan from a supplier, the NA sends the change to the PA and asks it to restart the direct plan trying to optimize resource utilization and satisfy as many customers. The PA transmits the result of the optimization model to the NA. The process is shown in figure 4

The solution proposed by the PA initiates the negotiation between the node and its customer. The negotiation is done in a round-trip message.

The negotiation process with customer is shown in figure 3

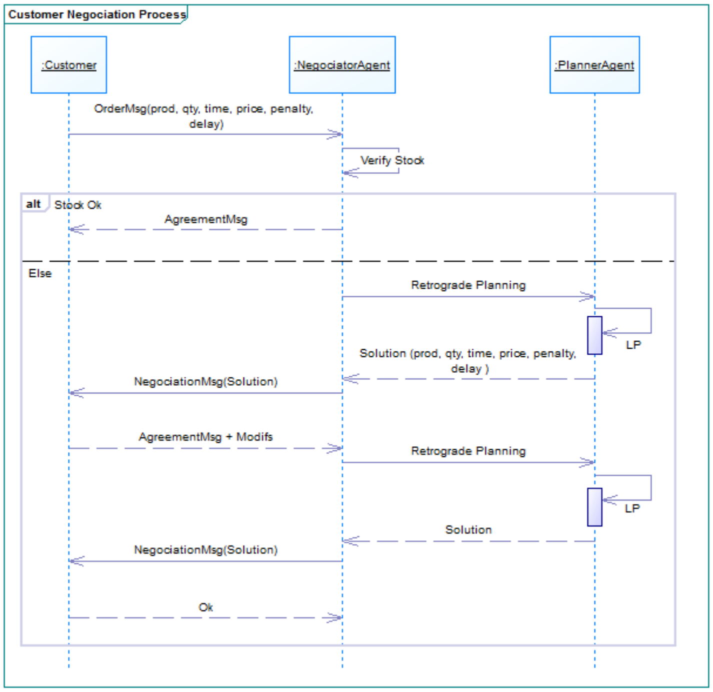


Figure 3. The process of negotiation between a node and custumer.

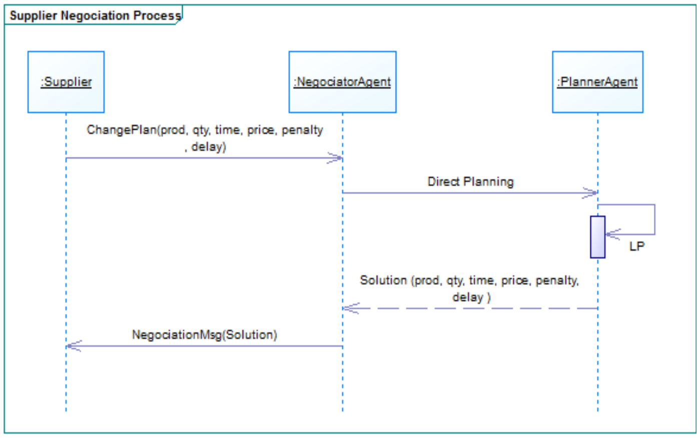

Supplier Negociation Process
:Supplier
:NegociatorAgent
:PlannerAgent
ChangePlan(prod, qty, time, price, penalty , delay)
Direct Planning
LP
Solution (prod, qty, time, price, penalty, delay)
NegociationMsg(Solution)

Figure 4. The process of negotiation between a node and supplier.

IV. CONCLUSION

Dynamic planning of the supply chain remains a difficult task because of the changes in production capacity and lead times. Thus, the success of each supply chain lies in the ability of partners to share information, synchronize their activities and try to work with a win-win principle to overcome contingencies. In this paper, we have opted for negotiation as a means of coordination that we have formalized using multi-agent systems.

REFERENCES

[1] Wang, M., Wang, H., Xu, D., Wan, K.K., Vogel, D.,"A web-service agent-based decision support system for securities exception management". Expert Systems with Applications 27 (3), 439–450, 2004

[2] Chiu, D.K.W., Yueh, Y.T.F., Leung, H.-f., Hung, P.C.K., "Towards ubiquitous tourist service coordination and process integration: a collaborative travel agent system with semantic web services." Information Systems Frontier, 10.1007/s10796-008-9087-2, 008.

[3] Parunak, H. VanDyke, "Manufacturing experience with the contract net."In M. Huhns, Distributed artificial intelligence, London/Los Altos, CA: Pitman/Morgan Kaufmann, pp. 285-310, 1987.

[4] Parunak, H. V. D., Baker, A. D., & Clark, S. J. "The aaria agent architecture: an example of requirements-driven agent-based system design." Agents-97, Marina del Rey, CA, 1997.

[5] Barbuceanu, M., & Fox, M.. "Capturing and modeling coordination knowledge for multi-agent systems." International Journal of Cooperative Information Systems, 5, pp 2-3, 1996. (http://www.ie.utoronto.ca/EIL/ABSpage/ABS-intro.html).

[6] Montreuil. B, J.M. Frayret, S. D'Amours, "A strategic framework for networked manufacturing", Computers in Industry, 42, pp 299–317, 2000.

[7] Lin, G. Y. -J., & Solberg, J. J. " Integrated shop floor control using autonomous agents". IIE Transactions: Design and Manufacturing, 24, 3, pp 57-71, 1992.

[8] Sikora, R., & Shaw, M. (1998). A multi-agent framework for the coordination and integration of information systems. Management Science, 40

[9] Cauvin, A.C.A., Ferrarini, A.F.A., Tranvouez, E.T.E., "Disruption management in distributed enterprise : A multi-agent modelling and simulation of cooperative recovery behaviours," international Journal of Production Economics, 122 (1), pp. 429-439, 2009

[10] L.A. Bearzotti, E.Salomone, O.J. Chiotti, "An autonomous multi-agent approach to supply chain event management," international Journal of Production Economics, 135 (1), pp. 468-478, 2012

[11] H.K. Chan, F.T.S. Chan, "Comparative study of adaptability and flexibility in distributed manufacturing supply chains", Decision Support Systems 48 (Issue 2), pp 331–341, 2010.

[12] Z. Guan, Application of decentralized cooperative problem solving in dynamic flexible scheduling, Proceedings of SPIE, vol. 2620, pp. 179–183, Bellingham, wach,1995

[13] S. Bussmann, J. Muller, "A negotiation framework for cooperating agents", Proceedings of CKBS-SIG, Dake Centre, University of Keele, UK, pp. 1-171, 992.

[14] S. Green, L. Hurst, N. Brenda, F. Somers, R. Evans, "Software agents: A review", Intelligent Agents, Group Report, 1997. http://www.cs.tcd.ie/research}groups/aig.iag/pubreview.zip.

A Study of Influential Factors in the Adoption and Diffusion of B2C E-Commerce

Rayed AlGhamdi (ال غامدي رائد د)
Faculty of Computing and IT
King Abdulaziz University
Jeddah, Saudi Arabia

Ann Nguyen
School of ICT, Griffith University
170 Kessels Rd, Nathan
QLD 4111, Australia

Vicki Jones
School of ICT, Griffith University
170 Kessels Rd, Nathan
QLD 4111, Australia

***Abstract*—This paper looks at the present standing of e-commerce in Saudi Arabia, as well as the challenges and strengths of Business to Customers (B2C) electronic commerce. Many studies have been conducted around the world in order to gain a better understanding of the demands, needs and effectiveness of online commerce. A study was undertaken to review the literature identifying the factors influencing the adoption and diffusion of B2C e-commerce. It found four distinct categories: businesses, customers, environmental and governmental support, which must all be considered when creating an e-commerce infrastructure. A concept matrix was used to provide a comparison of important factors in different parts of the world. The study found that e-commerce in Saudi Arabia was lacking in Governmental support as well as relevant involvement by both customers and retailers.**

Keywords-e-commerce;adoption; B2C; Saudi Arabia

I. INTRODUCTION

Saudi Arabia has a large and growing ICT marketplace. Yet, despite its size and rapid growth, progress in e-commerce activities is relatively slow [1-3]. The Saudi Government introduced e-commerce in 2001 in response to the fast expansion of e-commerce throughout the world. A permanent technical committee for e-commerce was established by the Saudi Ministry of Commerce. However, this Committee no longer exists, and from 2006, e-commerce supervision and development has been managed by the Ministry of Communications and Information Technology (MCIT). Unfortunately, there has been little progress since then [4].

II. FACTORS INFLUENCING THE ADOPTION AND DIFFUSION OF B2C E-COMMERCE

Around the world, many studies in online commerce have been conducted in order to gain a better understanding of its strengths and challenges. Research into the factors influencing the adoption and diffusion of B2C e-commerce tends to discuss these factors as belonging to one or more of four categories. These include: businesses, customers, environment and government facilitation.

A. *Factors influencing business' adoption of B2C e-commerce*

The literature discusses various issues that influence businesses to adopt B2C e-commerce. The highlighted issues in this review include

- organization e-readiness
- competitive pressure
- set-up and maintenance cost
- brand strength
- relative advantage of using e-commerce
- consumer purchasing power
- Privacy and Security
- type of products
- Resistance to change

B. *Factors influencing customers to purchase online*

The literature discusses various issues that influence consumers to purchase online. The highlighted issues in this review include

- lack of trust due to security/privacy concerns
- reluctance to use credit cards
- language barriers
- preferences for in-store shopping
- good quality of e-commerce websites
- lack of product trial / inspect by hand
- relative advantage (prices, convince etc)
- familiarity of products/seller's good reputation

C. *Environmental factors influencing the rate of B2C e-commerce adoption and diffusion*

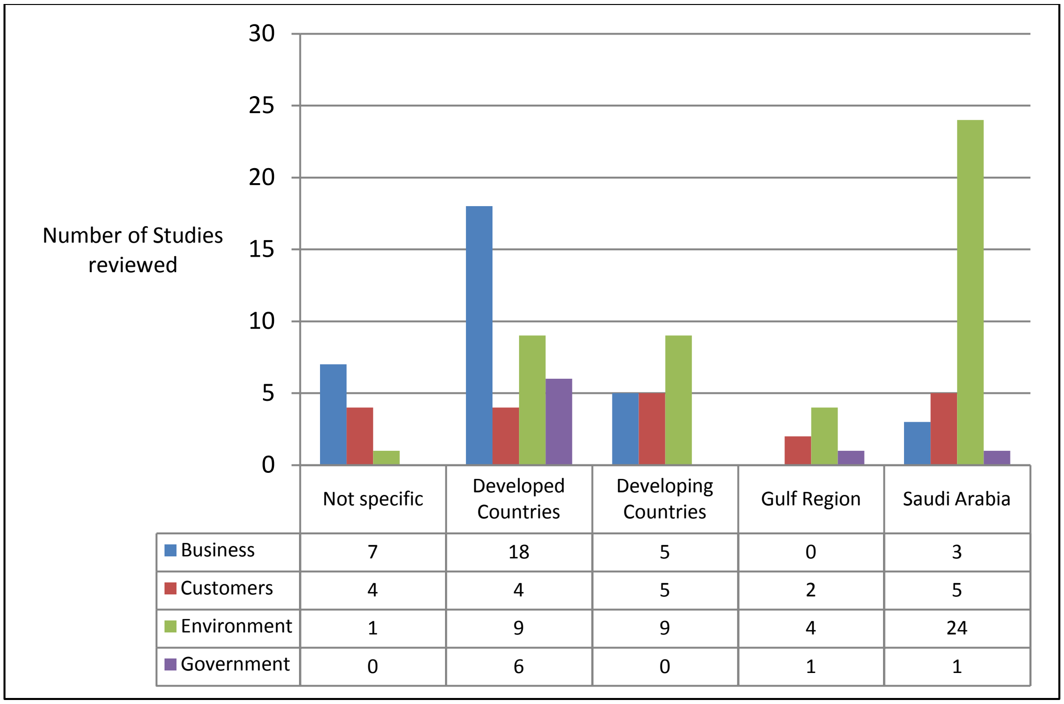
Environmental factors affect the online environment and e-commerce activities, and as a result, also affect businesses and customers. This means that these factors are influential when businesses choose to adopt e-commerce and when customers decide to start trading using e-commerce. The most highlighted issues include

- ICT infrastructure
- online payment mechanisms
- the degree of credit cards penetration
- legislative and regulatory framework
- logistics Infrastructure
- education and awareness

D. *Government intervention and its influence to the rate of B2C e-commerce adoption*

Throughout the World, governments tend to encourage and support the development of e-commerce. Also, as governments gain a larger Internet presents, their role often changes from simply informational to transactional. As a result, the government becomes both supplier and consumer, therefore contributing to the growth of e-commerce [5]. Using e-commerce also means that government departments can reach more customers, with faster service in a more economical way [6].

In Saudi Arabia, there are relatively few studies which identify business, consumers, and government factors. Most of the studies conducted, concentrate on environmental factors. Similarly with Developing Nations and the Gulf Region, the majority of studies have also focused on the environmental factors. However, studies conducted in Developed Countries tend to concentrate mainly on business and, to a lesser degree, government and environmental factors, with a smaller percentage on customers. This correlates with the fact that the environment, or e-commerce infrastructure, is already well established in these countries.



	Not specific	Developed Countries	Developing Countries	Gulf Region	Saudi Arabia
■ Business	7	18	5	0	3
■ Customers	4	4	5	2	5
■ Environment	1	9	9	4	24
■ Government	0	6	0	1	1

Figure 1. Factors and associated Studies which have influenced the adoption and diffusion of B2C e-commerce

The factors, which have influenced the adoption and diffusion of B2C e-commerce in research literature, are summarized in Table I (see the appendix). These factors and associated studies are further identified in Figure 1.

A study was undertaken to review the literature identifying the factors influencing the adoption and diffusion of B2C e-commerce. The review is divided into four sections: businesses, customers, environmental and governmental factors (see Table I, in the appendix).

A concept-centric structure is used to enable the separate influencing factors to be associated with the geographical context of the studies reviewed. The resulting concept matrix was then be used to provide a comparison of important factors in different parts of the world. The studies covered the geographical locations of: not specific, developed countries, developing countries, Gulf countries and Saudi Arabia. This division helps to compare different nations and identify similarities and differences (Figure 1).

In regards to Developing Nations, the Gulf Region and Saudi Arabia, the environment (e-commerce infrastructure) is not yet fully established, so is an important factor influencing the adoption and diffusion of B2C e-commerce. These figures also suggest that governmental support should be a high priority for e-commerce development. By contrast, the high level of readiness in e-commerce environment in Developed Countries leads most studies to concentrate on the businesses and why they might not be selling online when the e-commerce environment is ready for them.

III. E-GOVERNMENT IN SAUDI ARABIA

E-government and e-commerce share some similarity in terms of transaction requirements. Therefore, development in e-government can serve as an engine to power e-commerce development [5]. The similarity between e-government and e-commerce is that both of them depend on ICT infrastructure, online payment systems and mailing/post systems to reach their users/customers and deliver their services/products [7,8].

In 2003, a decision was taken by the Saudi Government to start work on e-government; however, a committee for e-government was established and the actual work started in 2005 [9,10]. With cooperation of three government entities (The Ministry of Communication and Information Technology (MCIT), the Ministry of Finance and the Communication and IT Committee (CITC)), an e-government program called ‘Yasser’ was launched in 2005 [7]. This program acts as an umbrella for all e-government activities, procedures, legislations and all related issues [10]. An e-government plan was set up with the following vision “By the end of 2010, everyone in the Kingdom will be able to enjoy from anywhere and at any time – world class government services offered in a seamless user friendly and secure way by utilizing a variety of electronic means" [7]. However, this vision has not been achieved as set up in a timely manner, which means the plan was not realistic [10]. The main problem, which was not taken into consideration, was the ICT infrastructure and assessing the e-readiness of the different government departments [11-14]. As a result, an e-government second action plan with the vision: "Enable use of efficient, integrated customer friendly and secure multiple e-Government services" (covering the period 2012-2016) has been launched; considering human resource training and development, promote cooperation and innovation culture, and maximizing efficiency of e-services provided by government agencies,[15].

IV. Government role in e-commerce promotion

Government support takes various forms from country to country; however, government regulation can be critical to supporting e-commerce [16]. Online shopping shows rapid growth in the developed world. Significantly, the South Korean government has played a key role promoting e-commerce. The Malaysian government is encouraging small and medium enterprises (SMEs) to adopt e-commerce solutions, and in Australia, the Government is providing support is various forms [17].

Although Australian online shoppers are increasing, many prominent Australian retailers are still lagging in terms of online sales. An Australian Government report about e-commerce stated that many retailers did not understand the potential benefits of online shopping and were concerned about the set-up and maintenance costs. The government followed by hosting an online retail forum to encourage, assist and inform retailers [17]. Small Business Online (SBO) was announced in the 2009-10 budget granting $14 million to help small businesses go online by offering training programs, advice and development of e-business resources [18].

The Government program, AusIndustry, supports business programs and a range of incentives aimed at helping businesses grow [18]. Singapore, South Korea and Hong Kong are also good examples of countries where the government has taken an active role in pushing for e-commerce proliferation [19]. The Singaporean government has been supporting e-commerce in the country since the early '90s. South Korea has a very strong ICT infrastructure, and Hong Kong set up and implemented an Electronic Transactions Ordinance in 2000. The Hong Kong ordinance contributed to future growth in e-commerce by providing for a legal infrastructure, such as the use digital signatures [20].

V. Discussion/Conclusion

From the study results, it seems that Saudi Arabia is in great need of more Governmental support. The other two areas which need to be increased are the customers and the retailers. Those countries with successful online retailing infrastructures have strong Governmental support. In Saudi Arabia, people tend to feel more confident in business ventures if they are backed by the Government [21]. It would seem logical that once the Government is involved, the element of trust that ensues can encourage customers and retailers to become involved.

A strategic plan needs to be developed to promote online retailing in Saudi Arabia. Based on this research it is apparent that e-commerce in Saudi Arabia is still in its early stages. With Government engagement, the current state of online retailing could transform into a fully integrated online retailing infrastructure. An appropriate method will be proposed in a later publication.

References

[1] S. Alfuraih, "E-commerce and E-commerce Fraud in Saudi Arabia: A Case Study," in *2nd International Conference on Information Security and Assurance* Busan, Korea, 2008, pp. 176-180.

[2] CITC (Communications and Information Technology Commission), "IT Report 2010 On the Internet Ecosystem in Saudi Arabia," Communications and Information Technology Commission, Riyadh2010.

[3] R. AlGhamdi, *et al.*, "Government Initiatives: The Missing Key for E-commerce Growth in KSA," in *International Conference on e-Commerce, e-Business and e-Service*, Paris, France, 2011, pp. 772-775.

[4] R. AlGhamdi and S. Drew, "Seven Key Drivers to Online Retailing in KSA," in *Proceedings of the IADIS International Conference on e-Society 2011*, Avila, Spain, 2011, pp. 237-244.

[5] C. J. Blakeley and J. H. Matsuura, "E-government: An engine to power e-commerce development," in *Proceedings of the European Conference on e-Government*, Dublin, Ireland, 2001, pp. 39-48.

[6] K. Layne and J. Lee, "Developing fully functional E-government: A four stage Model," *Government Information Quarterly 18*, pp.122–136, 2001.

[7] A. AL-Shehry, *et al.*, "The Motivations for Change Towards E-Government Adoption: Case Studies from Saudi Arabia," in *eGovernment Workshop*, London, UK, 2006.

[8] R. AlGhamdi, *et al.*, "Strategic Government Initiatives to Promote Diffusion of Online Retailing in Saudi Arabia," in *Sixth International Conference on Digital Information Management*, Melbourne, Australia, 2011, pp. 217-222

[9] I. Abu Nadi, L. Sanzogni, K. S. Sandhu and P. R.Woods, "Success Factors Contributing to eGovernment Adoption in Saudi Arabia: G2C approach", *Saudi International Innovation Conference SiiC 2008 Proceeding*, Leeds, UK, 2008, pp. 1-8

[10] O. Alfarraj, *et al.*, "eGovernment initiatives and key factors causing the delay of their implementation in Saudi Arabia," in *5th Conference on Qualitative Research in IT*, Brisbane, Australia, 2010, pp. 130-141.

[11] I. Abu Nadi, "Success Factors for eGovernment Adoption: Citizen Centric Approach", *LAP LAMBERT Academic Publishing,* Gold Coast, Australia, 2010.

[12] O. Alfarraj, *et al.*, "EGovernment Stage Model: Evaluating the Rate of Web Development Progress of Government Websites in Saudi Arabia," *International Journal of Advanced Computer Science and Applications (IJACSA),* vol. 2, pp. 82-90, 2011.

[13] M. Alshehri, *et al.*, "A Comprehensive Analysis of E-government services adoption in Saudi Arabia: Obstacles and Challenges," *International Journal of Advanced Computer Science and Applications (IJACSA),* vol. 3, pp. 1-6, 2012.

[14] M. Alshehri, S. Drew, T. Alhussain, and R. Alghamdi, "The Effects of Website Quality on Adoption of E-Government Service: AnEmpirical Study Applying UTAUT Model Using SEM", in J Lamp (ed.), 23rd Australasian Conference On Information Systems (ACIS 2012), Melbourne, Australia, pp. 1-13 (2012).

[15] Yasser eGov Program. (2012, 7 March). *The e-Government Second Action Plan (2012 – 2016)*. Available: http://www.yesser.gov.sa/en/MechanismsandRegulations/strategy/Pages/-second_Implementation_plan.aspx

[16] K. L. Kraemer, *et al.*, "Globalization and National Diversity: E-Commerce Diffusion and Impacts across Nations," in *Global e-commerce: impacts of national environment and policy*, K. L. Kraemer, *et al.*, Eds., ed New York: Cambridge Univ Press, 2006, pp. 13-61.

[17] DBCDE (Australian Department of Broadband. Communication and the Digital Economy). (2011, 14 May). *Online retail forum*. Available: http://www.dbcde.gov.au/digital_economy/online_retail_forum

[18] AusIndustry. (2009, 14 May). *Small Business Online (SBO)*. Available: http://www.ausindustry.gov.au/SmallBusiness/SmallBusinessOnline/Pages/SmallBusinessOnline.aspx

[19] M. Nair, "The E-commerce Ecology: Leapfrogging Strategies for Malaysia," in *ICT Strategic Review 2010/11 E-commerce for Global Reach* R. Ramasamy and S. Ng, Eds., ed Putrajaya, Malaysia: PIKOM (The National ICT Association of Malaysia), 2010, pp. 193-211.

[20] R. Wu, "Electronic Transactions Ordinance – Building a Legal Framework for E-commerce in Hong Kong," *The Journal of Information, Law and Technology (JILT)* http://www2.warwick.ac.uk /fac/soc/law/elj/jilt/2000_1/wu/

[21] R. AlGhamdi, S. Drew, T. Alhussain, " A Conceptual Framework for the Promotion of Trusted Online Retailing Environment in Saudi Arabia", *International Journal of Business and Management*, vol 7, no 5, pp. 140-149, 2012.

[22] S. Poon and P. Swatman, "An exploratory study of small business Internet commerce issues," *Information & Management,* vol. 35, pp. 9-18, 1999.

[23] D. Grewal, *et al.*, "Internet retailing: enablers, limiters and market consequences," *Journal of Business Research,* vol. 57, pp. 703-713, 2004.

[24] M. L. To and E. Ngai, "Predicting the organisational adoption of B2C e-commerce: an empirical study," *Industrial Management & Data Systems,* vol. 106, pp. 1133-1147, 2006.

[25] D. A. Colton, *et al.*, "Drivers of international e-tail performance: the complexities of orientations and resources," *Journal of International Marketing,* vol. 18, pp. 1-22, 2010.

[26] J. Kendall, *et al.*, "Electronic commerce adoption by SMEs in Singapore," in *Proceedings of the 34th Hawaii International Conference on System Sciences*, Hawaii 2001, pp. 1-10.

[27] P. B. Tigre, *et al.*, "Policies and Institutions for E-commerce Readiness: What can Developing Countries Learn from OECD Experience," 2002.

[28] J. Gibbs, *et al.*, "Environment and policy factors shaping global e-commerce diffusion: A cross-country comparison," *The Information Society,* vol. 19, pp. 5-18, 2003.

[29] K. V. Andersen, *et al.*, "Governance initiatives creating a demand-driven E-commerce approach: The case of denmark," *The Information Society,* vol. 19, pp. 95-105, 2003.

[30] S. A. Wymer and E. A. Regan, "Factors influencing e-commerce adoption and use by small and medium businesses," *Electronic Markets,* vol. 15, pp. 438-453, 2005.

[31] K. L. Kraemer, *et al.*, *Global e-commerce: impacts of national environment and policy*. New York: Cambridge Univ Press, 2006.

[32] S. C. Ho, *et al.*, "A growth theory perspective on B2C e-commerce growth in Europe: An exploratory study," *Electronic Commerce Research and Applications,* vol. 6, pp. 237-259, 2007.

[33] A. Scupola, "Government Intervention in SMEs' E-Commerce Adoption," in *Encyclopedia of Information Science and Technology*, M. Khosrow-Pour, Ed., Second ed: IGI Global, 2009, pp. 1689-1695.

[34] M. Tan and T. S. H. Teo, "Factors influencing the adoption of the Internet," *International Journal of Electronic Commerce,* vol. 2, pp. 5-18, 1998.

[35] A. M. Aladwani, "Key Internet characteristics and e-commerce issues in Arab countries," *Information and Management,* vol. 16, pp. 9-20, 2003.

[36] A. M. Sleem, "E-Commerce Infrastructure in Developing Countries," in *Electronic Business in Developing Countries: Opportunities and Challenges*, S. Kamel, Ed., ed USA: Idea Group Inc., 2006, pp. 349-385.

[37] Z. K. Shalhoub and S. L. AlQasimi, *The Diffusion of E-commerce in Developing Economies*. Cheltenham, UK: Edward Elgar Publishing Limited, 2006.

[38] M. M. Hafez, "The Role of Culture in Electronic Business Diffusion in Developing Countries," in *Electronic Business in Developing Countries: Opportunities and Challenges*, S. Kamel, Ed., ed USA: Idea Group Inc., 2006, pp. 34-44.

[39] G. R. El Said and G. H. Galal-Edeen, "The role of culture in e-commerce use for the Egyptian consumers," *Business Process Management Journal,* vol. 15, pp. 34-47, 2009.

[40] K. Al-Rawi, *et al.*, "Driving Factors for E-commerce: Gulf Region Review," *Academy of Information and Management Sciences Journal* vol. 11, pp. 19-32, 2008.

[41] K. W. Alrawi and K. A. Sabry, "E-commerce evolution: a Gulf region review," *International Journal of Business Information Systems,* vol. 4, pp. 509-526, 2009.

[42] B. H. Albadr, "E-commerce," *Science and Technology,* pp. 14-19 (Arabic source), 2003.

[43] S. M. Sait, *et al.*, "E-commerce in Saudi Arabia: Adoption and Perspectives," *Australasian Journal of Information Systems,* vol. 12, pp. 54-74, Sep 2004.

[44] A. Al-Solbi and P. J. Mayhew, "Measuring E-Readiness Assessment in Saudi Organisations Preliminary Results From A Survey Study," in *From e-government to m-government*, I. Kushchu and M. H. Kuscu, Eds., ed Brighton, UK: Mobile Government Consortium International LLC, 2005, pp. 467-475.

[45] S. Alwahaishi, *et al.*, "Electronic commerce growth in developing countries: Barriers and challenges " in *First International Conference on Networked Digital Technologies*, Ostrava, Czech Republic, 2009, pp. 225 - 232

[46] F. A. Aleid, *et al.*, "A consumers' perspective on E-commerce: practical solutions to encourage consumers' adoption of e-commerce in developing countries - A Saudi Arabian empirical study," in *International Conference on Advanced Management Science*, Chengdu, China, 2010, pp. 373-377.

[47] F. A. Aleid, *et al.*, "A suppliers' perspective on e-commerce: Suppliers responses to consumers' perspectives on e-commerce adoption in developing countries — A Saudi Arabian empirical study " in *Fifth International Conference on Digital Information Management (ICDIM)*, Thunder Bay, Canada, 2010, pp. 379-383.

[48] S. A. Al-Hudhaif and A. Alkubeyyer, "E-Commerce Adoption Factors in Saudi Arabia," *International Journal of Business and Management,* vol. 6, pp. 122-133, 2011.

[49] M. I. Eid, "Determinants of E-Commerce Customer Satisfaction, Trust, and Loyalty in Saudi Arabia," *Journal of Electronic Commerce Research,* vol. 12, pp. 78-93, 2011.

TABLE I. FACTORS WHICH HAVE INFLUENCED THE ADOPTION AND DIFFUSION OF B2C E-COMMERCE

Studies / Factors	Not specific				Developed Countries									Developing countries							Gulf Region		Saudi Arabia								
	[22]	[23]	[24]	[25]	[26]	[27]	[27]	[28]	[29]	[30]	[31]	[32]	[33]	[34]	[35]	[36]	[37]	[38]	[39]	[19]	[40]	[41]	[42]	[43]	[44]	[1]	[45]	[46]	[47]	[48]	[49]
Business																															
Organization e-readiness		●	●				●		●	●				●	●															●	
Competitive pressure			●				●			●				●																	
Set-up and maintenance cost										●																					
Brand strength				●																											
Relative advantage of using e-commerce	●		●		●		●	●		●	●			●															●		
Consumer purchasing power							●	●			●				●														●		
Privacy and Security										●	●																				
Type of products		●																													
Resistance to change							●	●																							
Customers																															
Lack of trust due to security/privacy concerns		●						●							●					●		●						●			●
Reluctance to use credit cards																	●										●	●			
Language barriers								●																							
Preferences for in-store shopping								●																							
Good quality of e-commerce websites		●						●														●									●
Lack of product trial / inspect by hand		●																													
Relative advantage (prices, convince etc)		●																													
Familiarity															●				●												

of products/seller’s + reputation																															
Environment																															
ICT infrastructure								●				●		●	●	●		●		●		●	●	●		●		●	●	●	
Online payment mechanisms								●								●	●					●				●	●			●	
The degree of credit cards penetration											●	●															●				
Legislative and regulatory framework									●		●					●						●	●		●			●	●	●	
Logistics Infrastructure		●																						●	●	●	●	●		●	
Education and Awareness									●			●								●		●		●			●		●		
Government																															
Government intervention/ promotion						●		●	●	●	●		●								●								●		

Comparative Study on Discrimination Methods for Identifying Dangerous Red Tide Species Based on Wavelet Utilized Classification Methods

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

***Abstract*—Comparative study on discrimination methods for identifying dangerous red tide species based on wavelet utilized classification methods is conducted. Through experiments, it is found that classification performance with the proposed wavelet derived shape information extracted from the microscopic view of the phytoplankton is effective for identifying dangerous red tide species among the other red tide species rather than the other conventional texture, color information.**

Keywords-hue feature; texture information; wavelet descripter; red tide; phytoplankton idintification

I. Introduction

Image retrieval success rate (search hit ratio) is not good enough due to a poor visual signature or image feature followed by a poor similarity measure as well as clustering and classification performance. There is some information which can be extracted from images. That is (1) Halftone, color, and spectral information, (2) Spatial information including shape, size, texture, etc., and (3) Relational information such as relation between objects and the other objects included in images. The conventional image retrieval methods use the color information such as HSV [1]: Hue, Saturation and Value (Intensity), RGB: Red, Green, and Blue, etc. as the spectral information. Meanwhile texture information is also used in conventional image retrieval methods as the spatial information. On the other hand, Bachattarian [1], Euclidian[2], Mahalanobis[3] distance measures [2] are well known as the similarity or distance measure. Not only hierarchical[4] and non-hierarchical clustering[5] as well as Bayesian rule of classification [6] and Maximum Likelihood classification[7], but also Vector quantization[8], Support vector machine [9], etc. are proposed and used for image retrievals. Relational information such as the relations among image portions or segments, semantic information, knowledge based information, relational similarity to classify semantic relations [3] etc. are tried to use in image retrievals. Spatial and spectral information derived from the images in concern is applicable image retrievals. There are some moment based spatial information extraction methods [4], [5], texture feature based spatial information extraction methods [6] and spectral information based image retrieval methods [7], [8], [9]. Furthermore, some attempts are made for image retrievals with wavelet descriptor as a spatial information extraction [9], [10]. In general, these conventional methods have not so good performance in terms of retrieval success rate.

All the spectral and spatial information are used in image retrieval except shape information. There are some trials to use shape information extracted from image using Fourier descriptor and the others. There are some definitions for Fourier descriptors. Zahn and Roskies proposed Z type descriptor [7] while Granlund proposed G type descriptor [11]. Z type descriptor is defined as the cumulative angle changes of the contour points from the starting point is expanded with Fourier series while G type descriptor defined as the length between the contour points from the start point of contour line in concern is expanded with Fourier series. Both of descriptors have the following problems:

(1) It is hard to express local properties,
(2) It cannot represent the shape of contour when the shape is not closed,

The results depend on the start point on the contour line in concern for tracking. On the other hand, Z type descriptor has another difficulty that the convergence speed is not fast so that it takes relatively large computational resources and the reproducibility of low frequency component is not good enough. Meanwhile, G type descriptor has another difficulty that Gibbs phenomenon [12] would occur at the end points of the closed curve of contour lines results in the end points cannot be preserved.

The shape descriptor proposed here is wavelet based descriptor not the Fourier type of descriptor. Therefore, the proposed wavelet based descriptor allows shape description through frequency-time analysis while Fourier based descriptor allows only frequency components representation of shape. There is some advantage for the wavelet based descriptor in shape information extraction rather than Fourier descriptor. Wavelet descriptor is proposed for best matching method to measure similarity between two feature vectors of the two shapes [9], [10]. This is impractical for higher dimensional feature matching. Therefore, wavelet descriptors are more suitable for model-based object recognition than

[1] http://www.cs.rit.edu/~ncs/color/t_convert.html
[2] http://en.wikipedia.org/wiki/Euclidean_distance
[3] http://en.wikipedia.org/wiki/Mahalanobis_distance
[4] http://en.wikipedia.org/wiki/Hierarchical_clustering
[5] http://www.daylight.com/meetings/mug96/barnard/E-MUG95.html
[6] http://en.wikipedia.org/wiki/Naive_Bayes_classifier
[7] http://www.ccrs.nrcan.gc.ca/glossary/index_e.php?id=341
[8] http://en.wikipedia.org/wiki/Learning_Vector_Quantization
[9] http://en.wikipedia.org/wiki/Support_vector_machine

data-driven shape retrieval, because for shape retrieval, which is usually conducted online, speed is essential.

Contour of the object extracted from the original image can be expressed with wavelet based descriptor. The proposed image retrieval method is based on the hue information and texture as well as the proposed wavelet described shape information of extracted objects to improve image retrieval success rate.

The following section describes the proposed image retrieval method followed by some experiments for reproducibility of the proposed wavelet descriptor in comparison to the conventional Fourier descriptor with several simple symmetrical and asymmetrical shapes. Then it is validated with the image database of phytoplankton [13].

II. Proposed Method

A. Research Background

There are a plenty of red tide species. Small portion of red tide species can be listed up in Figure 1.

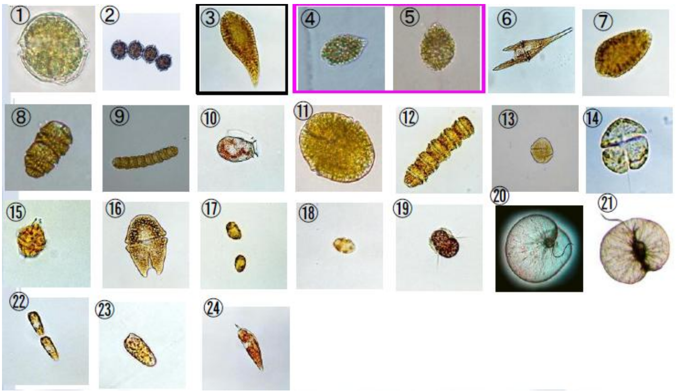
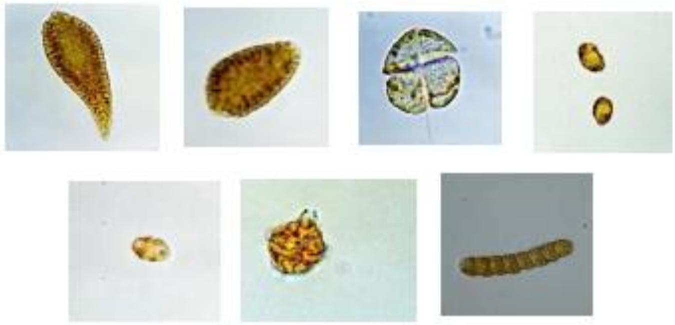
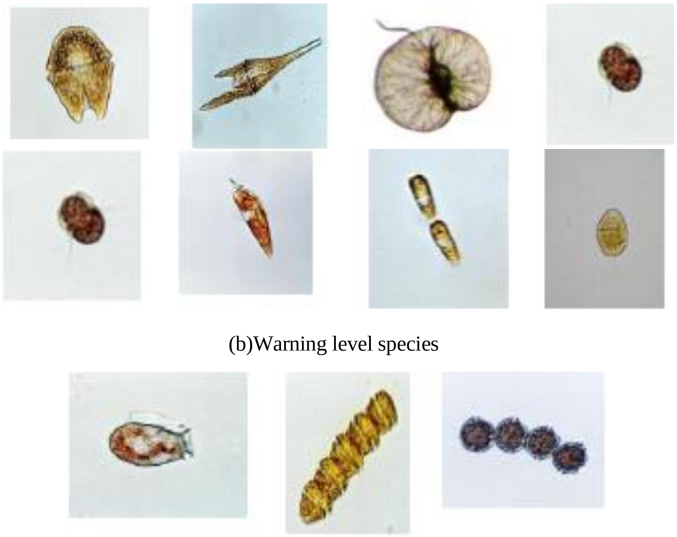


Figure 1 Photos of a portion of red tide species

These red tide species can be classified into three categories, (a) Caution level of species, (b) Warning level of species, and (c) Dangerous species. Fishes and shells take these dangerous red tide species. After that human habitats eat the fishes and shells. Then such persons get a bad situation and have an illness condition. Therefore, these red tide species are classified into dangerous species.



(a) Caution level species



(b)Warning level species

(c)Dangerous species

Figure 2 Three categories of red tide species

Identifying these dangerous red tide species is important. It, however, is not so easy to classify because these three categories of red tide species are quite resemble. Usually, the local fishery research institutes measure red tide from the research vessels with microscope. They used to count the number of red tide with microscope camera acquired imagery data on the ship. Then identify the red tide species in the same time quickly. Even though human perception capability is superior to that by machine learning based automatic classification, there are some mistakes. The purpose of the research is to improve classification performance by using considerable features which can be extracted from the microscopic imagery data.

B. Process Flow of the Proposed Image Classification

Image classification method based on hue information [14] and wavelet description based shape information [15] as well as texture information of the objects extracted with dyadic wavelet transformation [16] is proposed. Object is assumed to be focused so that the frequency component in the object is relatively high in comparison to the other (background). Figure 3 shows the process flow of the proposed image classification method.

One of the image features of hue information (angle) is calculated for the entire image in the color image database. Dyadic wavelet transformation[10] is also applied to the images then texture information is extracted from the transformed resultant image. Based on the Dyadic wavelet transformation, HH[11] image of edge is extracted from the original image. Morphological operations[12], opening and closing are then applied to the edge extracted images to remove inappropriate isolated pixels and undesirable image defects. After that the

[10] http://cas.ensmp.fr/~chaplais/Wavetour_presentation/ondelettes%20dyadiques/Dyadic_Transform.html

[11] HH denotes high frequency component in horizontal direction and high frequency component in vertical direction

[12] http://geol.hu/data/online_help/MorphologyFilters.html

resultant image is binarized with appropriate threshold then contour of the object is extracted. Then the Dyadic wavelet transformation is applied to the contour in order to extract shape information (Wavelet descriptor). After all, Euclidian distance between target image and the other candidate images in the color image database is calculated with extracted hue, texture and shape information then the closest image is retrieved.

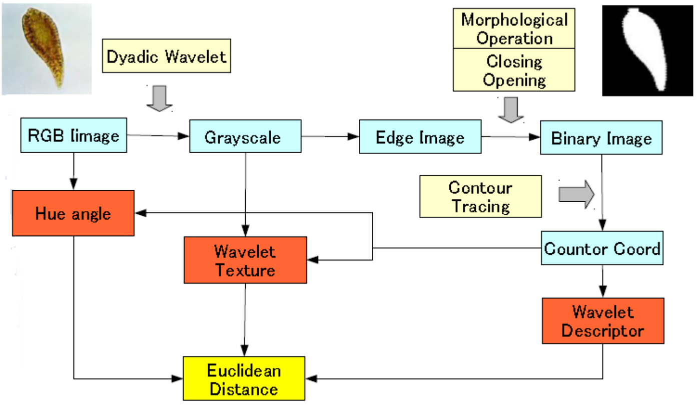


Figure 3 Process flow of the proposed image classification method.

C. Dyadic wavelet transformation

Using dyadic wavelet, frequency component can be detected. Dyadic wavelet allows to separate frequency components keeping image size with that of original image. Dyadic wavelet is called as a binary wavelet and has high pass and low pass filter components, {h[k], g[k]} and reconstruction filter {$\underline{h}$[k] ,$\underline{g}$[k]}. Low and high frequency components, C_n and d_n are expressed as follows,

$$C_n\,[i] = \Sigma_k\, h[k\,]C_{n-1}\,[i + k2^{n-1}\,] \qquad (1)$$
$$d_n\,[i] = \Sigma_k\, g\,[k\,]C_{n-1}\,[i + k2^{n-1}\,] \qquad (2)$$

Then original image is also reconstructed with the low and high frequency components as follows,

$$C_{n-1}\,[i] = 1/2\ \Sigma_k\, \underline{h}[k\,]C_n[i-k2^{n-1}] + \Sigma_k\, \underline{g}[k\,]d_n\,[i-k2^{n-1}\,] \quad (3)$$

If a new parameter s[m] is employed, then lifting dyadic wavelet is defined as follows,

$$h^{new}[k\,] = h^{old}\,[k\,] \qquad (4)$$
$$\underline{h}^{new}[k\,] = \underline{h}^{old}\,[\,k\,] + \Sigma_m\, s[-m]\,\underline{g}^{\,old}\,[k-m] \qquad (5)$$
$$g^{new}[k\,] = g^{old}\,[\,k\,] - \Sigma_m\, s[m]\,h^{old}\,[\,k-m] \qquad (6)$$
$$\underline{g}^{new}[k\,] = \underline{g}^{\,old}\,[k\,] \qquad (7)$$

D. Dyadic wavelet based descriptor (Shape information)

Image classification method with hue and texture information is conventional. In the proposed method, another feature, shape information is employed. Fourier descriptor is used, in general, to represent shape information. Although Fourier descriptor represents frequency component of the contour line, location information cannot be described. In other words, Fourier descriptor does support only frequency analysis, and does not support time-frequency component analysis.

Wavelet descriptor which is proposed by this paper supports a time-frequency component analysis so that not only frequency component but also location of contour edge can be discussed [17].

Let u(i) be distance between a point in the closed object contour line and a certain point i on the line, then the closed object contour line can be represented as u(i), i=1,2,…,n. i=1 corresponds to 0 degree while i=n corresponds to 360 degree, respectively. u(i) can be converted with dyadic wavelet transformation. Then the contour line can be represented with high frequency component of the dyadic wavelet transformed sequence as is shown in Figure 4. Then average of the high frequency component of pixel value is used for a feature of the image classification.

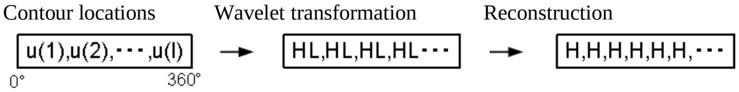


Figure 4 Dyadic wavelet descriptor for representation of the closed object contour lines.

E. Texture Information

Also texture information is useful for discrimination. Texture information can be derived from dyadic wavelet transformation. Texture information is defined as high frequency component of pixel value derived from dyadic wavelet transformation. Dyadic wavelet transformation is applied to the 2x2 pixels defined in Figure 5. Pixel value of the pixel in the object is replaced to the high frequency component detected with dyadic wavelet. Thus image which represents texture information of the detected object image is generated [18].

F. Hue angle

Thus contour of the object is detected. Then Red, Green, and Blue: RGB of the original object image can be transformed to Hue, Saturation, and Intensity: HSV information. Hue information in unit of radian, in particular, is useful for discrimination of the target image classifications of phytoplankton images. HSV, on the other hand, is expressed in Figure 6 (Color coordinate system).

RGB to HSV conversion is also be expressed as follows,

$$V = \max(R, G, B)$$
$$S = (V - X)/V \text{ where } X = \min(R, G, B)$$
$$R = V\!: H = (\pi/3)\ (b - g)$$
$$G = V\!: H = (\pi/3)\ (2 + r - b)$$
$$B = V\!: H = (\pi/3)\ (4 + g - r)$$

where $r = (V - R)/(V - X)$, $g = (V - G)/(V - X)$, $b = (V - B)/(V - X\ )$, H ranges from 0 to 360, S ranges from 0 to 1, V ranges from 0 to 1, HSV representation and R, G, B also range from 0 to 1. These three features, hue, *H*, texture, *xx* and shape information, *yy* composes three dimensional feature space results in measurement of Euclidian distance between a query image and the images in previously created image database. Using the distance, a query image can be retrieved from the image in the database. Thus image classifications can be done with hue and texture information as well as shape information derived from dyadic wavelet descriptor.

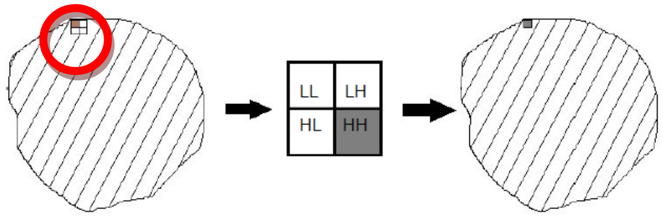


Figure 5. Detected object and 2x2 of matrix in the object to detect texture information with 2x2 of dyadic wavelet transformation.

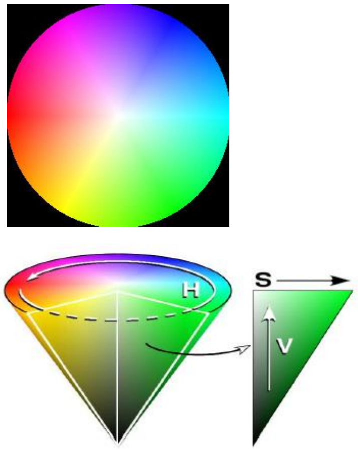


Figure 6 Hue angle and Saturation as well as Value (Intensity).

G. Alternative shape information (Fourier Descriptor)

The location coordinate is expressed in the complex plane representation for the G type of Fourier descriptor, that is,

$$Z_s = X_s + iY_s \qquad (11)$$

Then space or time domain locations can be transformed with eth following equation,

$$F_t = \frac{1}{S}\sum_{s=0}^{S-1} Z_s \exp\left(\frac{-2\pi ks}{S} i\right) \qquad (12)$$

It can be inversely transformed with the following equation,

$$Z_s = \sum_{t=0}^{S-1} F_t \exp\left(\frac{2\pi ks}{S} i\right) \qquad (13)$$

If these transformation and inverse transformation is perfect, then the original shapes are completely reproduced. The reproducibility for the shapes of circle, triangle, square, and trapezium (asymmetric shape) of the proposed wavelet descriptor is better than that of the conventional Fourier descriptor as shown in Figure 7.

In the comparison, the original image is binarized and the contour is extracted. Then shape information is extracted with both Fourier descriptor (G-type) and Dyadic wavelet descriptor. After that, image is reconstructed with the extracted shape information then compares the reconstructed images with two descriptors, Fourier and Dyadic wavelet descriptors. The difference between the reconstructed contours and original image is shown in Table 1. Thus it is found that the reproducibility of Dyadic wavelet descriptor is better than the conventional Fourier descriptor.

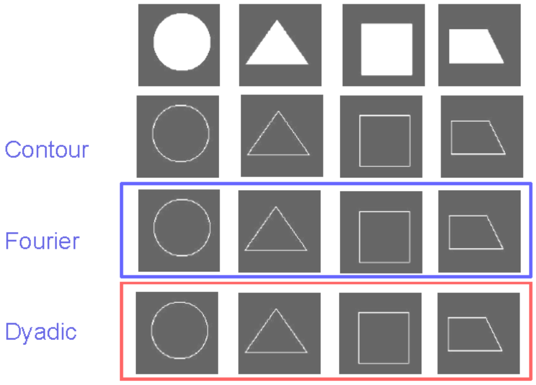


Figure 7 Comparison of reproducibility of the shapes between Fourier and Dyadic wavelet descriptor.

TABLE I. COMPARISON OF THE DIFFERENCE BETWEEN THE ORIGINAL AND RECONSTRUCTED CONTOURS WITH FOURIER AND DYADIC WAVELET DESCRIPTORS.

Shape	Fourier	Dyadic
circle	0.4121	0.1809
triangle	0.5391	0.1280
square	0.3689	0.1101
trapezium	0.4660	0.1929

This method can be expanded to 3D object. Once 3D object image is acquired through scanning in roll/pitch/yaw directions with the appropriate step angle, and then contour lines of the acquired 2D images are extracted. After that the 3D object shape complexity is represented with the wavelet descriptor as a resultant image which includes series of the high frequency components derived from dyadic wavelet transformation as shown in Figure 4 something like the following sequence,

(1) HH..H(for roll angle 0), HH..H(for roll angle 10),.., HH..H(for roll angle 350),
(2) HH..H(for pitch angle 0), HH..H(for pitch angle 10),.., HH..H(for pitch angle 350),
(3) HH..H(for yaw angle 0), HH..H(for yaw angle 10),.., HH..H(for yaw angle 350).

It is an image and is totally visual. This image represents 3D object shape complexity as an index. Also this index is shift invariant and rotation invariant. Namely, the index is not changed even if 3D object is translated and rotated.

III. EXPERIMENTS

A. Euclidian Distance

One of the measures for classification performance evaluation is Euclidian distance among the classes in concern. Shorter Euclid distance implies a poor classification performance while longer distance means a good performance. If the texture and hue information are used for classification, then Table 2 of Euclid distance is calculated while those of the case of utilizing wavelet descriptor and texture information is shown in Table 3. In Table 2 and 3, Euclid distance between class #5 of *Chattnella_Antiqua* and the others are listed because primary red tide species at this moment is *Chattnella_Antiqua*.

TABLE II. CALCULATED EUCLID DISTANCE WHEN TEXTURE AND HUE INFORMATION ARE USED AS FEATURES

		Euclid Distance
	file name	（t、h）
1	a.catenella1cell	0.0342
2	a.catenella4cell	0.7316
3	c.antiqua	0.0024
4	c.antiqua2	0.0602
5	c.antiqua3	0.0000
6	c.furca	0.0798
7	c.marina	0.0074
8	c.polykrikoides2cell	0.0035
9	c.polykrikoides8cell	0.0148
10	d.fortii	0.0817
11	g.catenatum1cell	0.0061
12	g.catenatum5cell	0.0012
13	g.instriatum	0.0470
14	g.mikimotoi	0.0448
15	g.polygramma	0.0214
16	g.sanguineum	0.0356
17	h.akashiwo	0.0260
18	h.circularisquama	0.0148
19	m.rubrum	0.0044
20	n.scintillans4	0.2059
21	n.scintillans5	0.7316
22	p.dentatum	0.0196
23	p.dentatum2	0.0358
24	p.signoides	0.0271

TABLE III. CALCULATED EUCLID DISTANCE WHEN TEXTURE AND WAVELET DESCRIPTER INFORMATION ARE USED AS FEATURES

	Euclid Distance（w、t）		
	daubechies1	dyadic	fourier
1	0.0044	0.0263	0.0234
2	0.7619	0.6435	0.5593
3	0.0034	0.0082	0.0565
4	0.0104	0.0886	0.0211
5	0.0000	0.0000	0.0000
6	0.0071	0.0048	0.0719
7	0.0026	0.0281	0.0061
8	0.0136	0.0329	0.0323
9	0.0259	0.0436	0.0279
10	0.0098	0.0679	0.0104
11	0.0032	0.0730	0.0186
12	0.0144	0.0419	0.0385
13	0.0021	0.0488	0.0192
14	0.0114	0.0405	0.4726
15	0.0158	0.0305	0.0142
16	0.0053	0.0319	0.0189
17	0.0148	0.0136	0.0063
18	0.0016	0.1090	0.0287
19	0.0164	0.0294	0.3365
20	0.0147	0.1654	0.0651
21	0.7619	0.6435	0.5593
22	0.0087	0.0119	0.0411
23	0.0260	0.0025	0.0407
24	0.0086	0.0356	0.0650

Also, Euclid distance between *Chattnella_Antiqua* and the other species are calculated and shown in Table 4 for the case of utilizing all these three features of wavelet descriptor, texture, and hue information together.

TABLE IV. CALCULATED EUCLID DISTANCE WHEN TEXTURE WAVELET DESCRIPTER AND HUE INFORMATION ARE USED AS FEATURES

	Euclid Distance（w、t、h）		
	daubechies1	dyadic	fourier
1.0000	0.0345	0.0431	0.0414
2.0000	0.9294	0.8352	0.7722
3.0000	0.0042	0.0086	0.0565
4.0000	0.0611	0.1072	0.0638
5.0000	0.0000	0.0000	0.0000
6.0000	0.0800	0.0799	0.1074
7.0000	0.0078	0.0291	0.0096
8.0000	0.0141	0.0330	0.0325
9.0000	0.0298	0.0460	0.0316
10.0000	0.0823	0.1062	0.0823
11.0000	0.0069	0.0732	0.0195
12.0000	0.0144	0.0419	0.0385
13.0000	0.0471	0.0678	0.0508
14.0000	0.0462	0.0604	0.4747
15.0000	0.0266	0.0373	0.0257
16.0000	0.0360	0.0478	0.0403
17.0000	0.0299	0.0293	0.0267
18.0000	0.0149	0.1099	0.0323
19.0000	0.0170	0.0297	0.3365
20.0000	0.2064	0.2641	0.2159
21.0000	0.9294	0.8352	0.7722
22.0000	0.0214	0.0229	0.0455
23.0000	0.0442	0.0359	0.0542
24.0000	0.0284	0.0448	0.0704

IV. CONCLUSION

Comparative study on discrimination methods for identifying dangerous red tide species based on wavelet utilized classification methods is conducted. Through experiments, it is found that classification performance with the proposed wavelet derived shape information extracted from the microscopic view of the phytoplankton is effective for identifying dangerous red tide species among the other red tide species rather than the other conventional texture, color information

ACKNOWLEDGMENT

The author would like to thank Mr. Yuji Yamada for his effort to conduct the experiments.

REFERENCES

[1] Duda R.O., P.E. Hart, and D.G. Stork, (2001), Pattern Classification, (Second Edition), John Wiley & Sons Inc.

[2] Arai K. (1996), Fundamental theory for image processing, Gakujutsu-Tosho Shuppan Publishing Co., Ltd.

[3] Séaghdha, D.O., Ann Copestake, (2009), Using lexical and relational similarity to classify semantic relations, Computational Linguistics 621-629.

[4] Teh C.H. and R. T. Chin,(1988), On image analysis by the methods of moments, IEEE Trans. On Pattern Analysis and Machine Intelligence, 10, 4, 496-513

[5] Taubin G. and D. B. Cooper,(1991), Recognition and Positioning of Rigid Objects Using Algebraic Moment Invariants, SPIE Conf. On Geometric Methods in Computer Vision, 1570, 175-186.

[6] Niblack W.,(1993), The QBIC Project: Querying Images By Content Using Color, Texture and Shape, SPIE Conf. On Storage and Retrieval for Image and Video Databases, 1908, 173-187.

[7] Zahn C.T., and Ralph Z. Roskies. (1972), Fourier Descriptors for Plane closed Curves. IEEE Trans. On Computer,c-21(3):269-281.

[8] Huang C.L. and D.H. Huang,(1998), A Content-based image retrieval system. Image and Vision Computing, 16:149-163.

[9] Yang H.S., S.U. Lee, K M. Lee., (1998), Recognition of 2D Object Contours Using Starting-Point-Independent Wavelet Coefficient Matching. Journal of Visual Communication and Image Representation, 9, 2, 171-181.

[10] Tieng Q.M. and W. W. Boles, (1997), Recognition of 2D Object Contours Using the Wavelet Transform Zero-Crossing Representation, IEEE Trans. on PAMI 19, 8, 1997.

[11] Grandlund H., (1972), Fourier preprocessing for hand print character recognition, IEEE Trans. on Computers, 621, 195-201.

[12] Gibbs, J. W., (1899), "Fourier Series". Nature 59, 200 and 606.

[13] Arai K. and Yasunori Terayama (2010), Polarized radiance from red tide, Proceedings of the SPIE Asia Pacific Remote Sensing, AE10-AE101-14.

[14] Arai K. et al. (1991), Takagi and Shimoda edt., Image Analysis Handbook, Tokyo Daigaku Shuppan-kai publishing.

[15] Arai K. (1998), Methods for Image Processing and Analysis of Earth Observation Satellite Imagery Data, Morikita Shuppan Publishing Co., Ltd.

[16] Arai K. and L. Jameson (2001), Earth observation satellite data analysis based on wavelet analysis, Morikita-Shuppan Publishing Co., Ltd.

[17] Arai K. (2002), Java based Earth observation satellite imagery data processing and analysis, Morikita-Shuppan Publishing Co., Ltd.

[18] Arai, K., (2011), Visualization of 3D object shape complexity with wavelet descriptor and its application to image retrievals, Journal of Visualization, 15, 155-166.

[19] Datta, Ritendra; Dhiraj Joshi, Jia Li, James Z. Wang (2008). "Image Retrieval: Ideas, Influences, and Trends of the New Age". *ACM Computing Surveys* 40 (2): 1–60

[20] Prasad, B E; A Gupta, H-M Toong, S.E. Madnick (1987). "A microcomputer-based image database management system". IEEE Transactions on Industrial Electronics IE-34 (1): 83–8.

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

APPENDIX

Table 1, 2, and 3 shows the Euclid distance between *Chattnella_Antiqua* and the other species for the case that the support length of base function of wavelet is two. Meanwhile, Table 4, 5, and 6 shows the Euclid distance between *Chattnella_Antiqua* and the other species for the case that the support length of base function of wavelet is four.

		Euclid Distance
	file name	（t、h）
1	a.catenella1cell	0.0342
2	a.catenella4cell	0.7316
3	c.antiqua	0.0024
4	c.antiqua2	0.0602
5	c.antiqua3	0.0000
6	c.furca	0.0798
7	c.marina	0.0074
8	c.polykrikoides2cell	0.0035
9	c.polykrikoides8cell	0.0148
10	d.fortii	0.0817
11	g.catenatum1cell	0.0061
12	g.catenatum5cell	0.0012
13	g.instriatum	0.0470
14	g.mikimotoi	0.0448
15	g.polygramma	0.0214
16	g.sanguineum	0.0356
17	h.akashiwo	0.0260
18	h.circularisquama	0.0148
19	m.rubrum	0.0044
20	n.scintillans4	0.2059
21	n.scintillans5	0.7316
22	p.dentatum	0.0196
23	p.dentatum2	0.0358
24	p.signoides	0.0271

	Euclid Distance（w、t）		
	daubechies1	dyadic	fourier
1	0.0044	0.0103	0.0234
2	0.7619	0.6027	0.5593
3	0.0034	0.0141	0.0565
4	0.0104	0.0775	0.0211
5	0.0000	0.0000	0.0000
6	0.0071	0.0064	0.0719
7	0.0026	0.0146	0.0061
8	0.0136	0.0264	0.0323
9	0.0259	0.0303	0.0279
10	0.0098	0.0510	0.0104
11	0.0032	0.0508	0.0186
12	0.0144	0.0407	0.0385

13	0.0021	0.0287	0.0192
14	0.0114	0.0345	0.4726
15	0.0158	0.0297	0.0142
16	0.0053	0.0185	0.0189
17	0.0148	0.0041	0.0063
18	0.0016	0.0892	0.0287
19	0.0164	0.0169	0.3365
20	0.0147	0.1305	0.0651
21	0.7619	0.6027	0.5593
22	0.0087	0.0250	0.0411
23	0.0260	0.0250	0.0407
24	0.0086	0.0455	0.0650

	Euclid Distance（w、t、h）		
	daubechies1	dyadic	fourier
1	0.0345	0.0357	0.0414
2	0.9294	0.8041	0.7722
3	0.0042	0.0143	0.0565
4	0.0611	0.0982	0.0638
5	0.0000	0.0000	0.0000
6	0.0800	0.0800	0.1074
7	0.0078	0.0163	0.0096
8	0.0141	0.0267	0.0325
9	0.0298	0.0337	0.0316
10	0.0823	0.0963	0.0823
11	0.0069	0.0511	0.0195
12	0.0144	0.0407	0.0385
13	0.0471	0.0551	0.0508
14	0.0462	0.0566	0.4747
15	0.0266	0.0366	0.0257
16	0.0360	0.0401	0.0403
17	0.0299	0.0263	0.0267
18	0.0149	0.0904	0.0323
19	0.0170	0.0174	0.3365
20	0.2064	0.2438	0.2159
21	0.9294	0.8041	0.7722
22	0.0214	0.0318	0.0455
23	0.0442	0.0437	0.0542
24	0.0284	0.0529	0.0704

		Euclid Distance
	file name	（t、h）
1	a.catenella1cell	0.0342
2	a.catenella4cell	0.7316
3	c.antiqua	0.0024
4	c.antiqua2	0.0602
5	c.antiqua3	0.0000
6	c.furca	0.0798
7	c.marina	0.0074
8	c.polykrikoides2cell	0.0035
9	c.polykrikoides8cell	0.0148
10	d.fortii	0.0817
11	g.catenatum1cell	0.0061
12	g.catenatum5cell	0.0012
13	g.instriatum	0.0470
14	g.mikimotoi	0.0448
15	g.polygramma	0.0214
16	g.sanguineum	0.0356
17	h.akashiwo	0.0260
18	h.circularisquama	0.0148
19	m.rubrum	0.0044
20	n.scintillans4	0.2059
21	n.scintillans5	0.7316
22	p.dentatum	0.0196
23	p.dentatum2	0.0358
24	p.signoides	0.0271

	Euclid Distance（w、t）		
	daubechies1	dyadic	fourier
1	0.0044	0.0097	0.0234
2	0.7619	0.5513	0.5593
3	0.0034	0.0161	0.0565
4	0.0104	0.0595	0.0211
5	0.0000	0.0000	0.0000
6	0.0071	0.0186	0.0719
7	0.0026	0.0029	0.0061
8	0.0136	0.0126	0.0323
9	0.0259	0.0126	0.0279
10	0.0098	0.0283	0.0104
11	0.0032	0.0222	0.0186
12	0.0144	0.0373	0.0385
13	0.0021	0.0007	0.0192
14	0.0114	0.0307	0.4726
15	0.0158	0.0268	0.0142
16	0.0053	0.0008	0.0189
17	0.0148	0.0114	0.0063
18	0.0016	0.0630	0.0287
19	0.0164	0.0088	0.3365
20	0.0147	0.0830	0.0651
21	0.7619	0.5513	0.5593
22	0.0087	0.0356	0.0411
23	0.0260	0.0293	0.0407
24	0.0086	0.0455	0.0650

	Euclid Distance（w、t、h）		
	daubechies1	dyadic	fourier
1	0.0345	0.0355	0.0414
2	0.9294	0.7664	0.7722
3	0.0042	0.0163	0.0565
4	0.0611	0.0847	0.0638
5	0.0000	0.0000	0.0000
6	0.0800	0.0819	0.1074
7	0.0078	0.0079	0.0096
8	0.0141	0.0131	0.0325
9	0.0298	0.0194	0.0316
10	0.0823	0.0864	0.0823
11	0.0069	0.0230	0.0195
12	0.0144	0.0373	0.0385
13	0.0471	0.0470	0.0508
14	0.0462	0.0543	0.4747
15	0.0266	0.0343	0.0257
16	0.0360	0.0356	0.0403
17	0.0299	0.0283	0.0267
18	0.0149	0.0647	0.0323
19	0.0170	0.0098	0.3365
20	0.2064	0.2220	0.2159
21	0.9294	0.7664	0.7722
22	0.0214	0.0406	0.0455
23	0.0442	0.0463	0.0542
24	0.0284	0.0530	0.0704

Nonlinear Mixing Model of Mixed Pixels in Remote Sensing Satellite Images Taking Into Account Landscape

Verification of the proposed nonlinear pixed pixel model through simulation studies

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

***Abstract*— Nonlinear mixing model of mixed pixels in remote sensing satellite images taking into account landscape is proposed. Most of linear mixing models of mixed pixels do not work so well because the mixed pixels consist of several ground cover targets in a nonlinear basis essentially. In particular, mixing model should be nonlinear because reflected photons from a ground cover target are scattered with atmospheric continuants and then reflected by the other or same ground cover targets. Therefore, mixing model has to be nonlinear. Monte Carlo Ray Tracing based nonlinear mixing model is proposed and simulated. Simulation results show a validity of the proposed nonlinear mixed pixel model.**

Keywords— nonlinearity; mixed pixels; Monte Carlo Ray Tracing; landscape

I. Introduction

Most of linear mixing models of mixed pixels do not work so well because the mixed pixels consist of several ground cover targets in a nonlinear basis essentially. In particular, mixing model should be nonlinear because reflected photons from a ground cover target are scattered among the ground cover targets and with the atmospheric continuants and then reflected by the other or same ground cover targets. Therefore, mixing model has to be nonlinear essentially. Also nonlinear mixing model has to take into account landscape.

The pixels in earth observed images which are acquired with Visible to Near Infrared: VNIR sensors onboard remote sensing satellites are, essentially mixed pixels (mixels) which consists of several ground cover materials [1]. Some mixel model is required for analysis such as un-mixing of the mixel in concern [2],[3]. Typical mixel is linear mixing model which is represented by linear combination of several ground cover materials with mixing ratio for each material [4]. It is not always true that the linear mixel model is appropriate [5]. Due to the influences from multiple reflections between the atmosphere and ground, multiple scattering in the atmosphere on the observed radiance from the ground surface, pixel mixture model is essentially non-linear rather than linear. These influence is interpreted as adjacency effect [6].[7].

Landscape based nonlinear mixing model of the mixed pixels of remote sensing satellite images which takes into account scattering due to the atmospheric molecules and aerosol particles in the atmosphere is proposed. The proposed model is based on the well-known Monte Carlo Ray Tracing: MCRT model [8]. Simulation cell which is composed with the atmosphere, extrat errestrial solar irradiance and ground surface is assumed. Landscape is modeled with the ground surface cells by referencing to Digital Elevation Model: DEM which is derived from stereo pair of visible to near infrared radiometers onboard remote sensing satellites.

In the atmosphere, there are Rayleigh and Mie scatterings due to the atmospheric molecules and aerosol particles, respectively. Therefore, refractive index and size distribution has to be determined or designated for aerosol particles. The proposed nonlinear mixing model is primarily for analyzing mixed pixels: Mixels which are acquired with visible to near infrared radiometer. Therefore, water vapor, ozone (and of course, oxygen and nitrogen gasses) is assumed to be the atmospheric continuants because these atmospheric continuants have absorption in the wavelength region. The proposed model is realized in a computer simulation and is validated.

The following section describes the proposed nonlinear mixing model followed by a method for computer simulation method. Then the proposed nonlinear model is validated with MCRT model based computer simulation. Finally, conclusion is described together with some discussions.

II. Proposed Nonlinear Mixing Model Of Mixed Pixels In Remote Sensing Satellite Images

A. *Monte Carlo Ray Tracing Simulation Model*

Figure 1 shows simulation cell of MCRT model for the proposed nonlinear mixing model of Mixels of visible to near infrared radiometer acquired remote sensing satellite images. The simulation cell size is 50 km by 50 km by 50 km. The ground surface parameters are reflectance and elevation while those of the atmosphere are optical depth of the atmospheric molecules and aerosol particles.

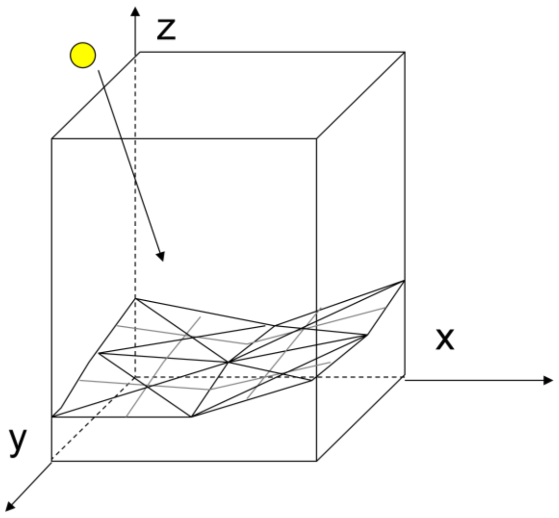


Figure 1. Simulation cell for Monte Carlo Ray Tracing

MCRT process flow is shown in Figure 2. 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. 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 [1] while that by aerosol is followed by Mie scattering law [1]. On the other hands, the photons which reach on the ground are reflected and absorbed at the surface depending on the surface reflectance.

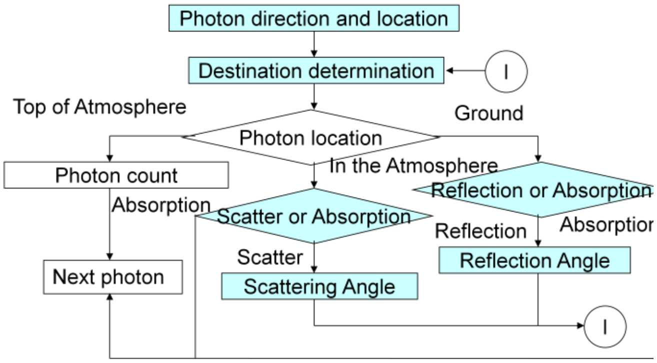


Figure 2. MCRT process flow.

B. Ground Surface Model

Ground cover targets are situated with their own reflectance and slopes. One pixel, Instantaneous Field of View: IFOV can be divided into four by four sub-pixels as shown in Figure 3.

Simulation cell size is 50 km cube. Therefore, one sub-pixel is composed with 1 km by 1 km (50 by 50 sub-pixels in the simulation cell). Also the numbers in the sub-pixel, Ref[a][b] denotes the reflectance of the ground surface. Elevation at the four corners of a pixel can be derived from DEM which is obtained with stereo pair of visible to near infrared radiometer imagery data. Elevation at the center of the pixel in concern can be estimated with the elevation data at the neighboring four corners as shown in Figure 4. Thus the slopes of the triangles which are shown in Figure 4 are determined.

Three points of the triangle are known. Therefore, equations for representation of triangles are known as shown in Figure 5. Also one pixel is composed with four by four sub-pixels and consists of 16 of sub-surface slopes as shown in Figure 6. Elevations of the four corners of the sub-pixel are given by DEM which is derived from stereo pair of images of visible to near infrared radiometer onboard remote sensing satellites. Therefore, slopes are calculated for all the 16 of sub-surface of pixel.

Ref[0][0]	Ref[0][1]	Ref[0][2]	Ref[0][3]
Ref[1][0]	Ref[1][1]	Ref[1][2]	Ref[1][3]
Ref[2][0]	Ref[2][1]	Ref[2][2]	Ref[2][3]
Ref[3][0]	Ref[3][1]	Ref[3][2]	Ref[3][3]

Figure 3. Visible to near infrared radiometer acquired pixel is assumed to be combined four by four sub-pixels.

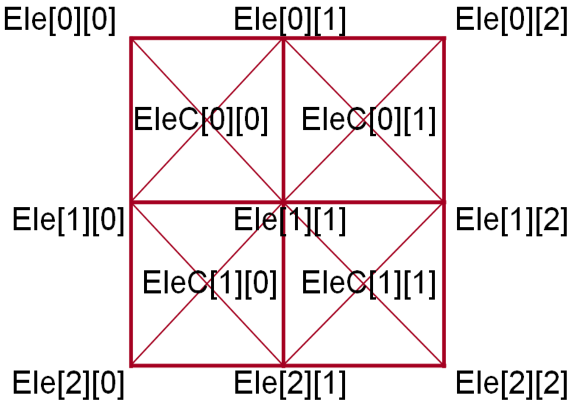


Figure 4. Landscape model of the pixel in concern.

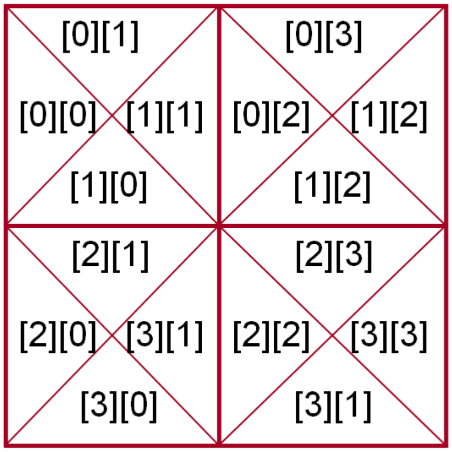


Figure 5. Calculation of triangle equation of the pixel in concern

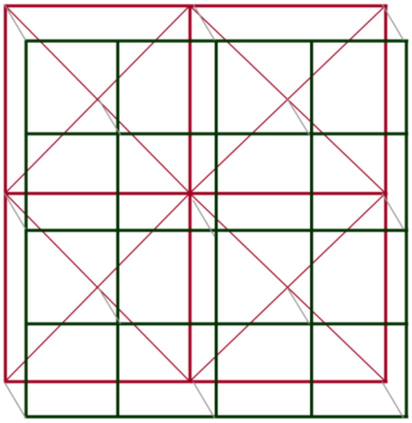


Figure 6. Pixel which is composed with four by four sub-pixels and consists of 16 of sub-surface slopes.

III. EXPERIMENTS

A. Preliminary Simulation Study

Figure 7 shows ground surface model. Algorithm for determination of elevation is as follows,

- A+B=20 , A>B
- Average elevation=10
- A=10,11,…,19,20
- B=10, 9 ,…, 1 ,0
- Standard deviation=(A -10)× √2

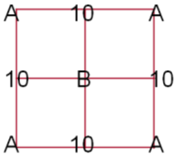


Figure 7. GROUND surface model used for simulation study

Thus the mean elevation of the surface in the simulation cell is 10 and standard deviation of elevation square root 2 (0.5).

There are four types of photon passes as shown in Figure 8,

(1) Photons are scattered by atmospheric molecules and aerosol particles and are come out from the top of the atmosphere,

(2) Photons are scattered by atmospheric molecules and aerosol particles and are absorbed in the atmosphere,

(3) Photons are scattered by atmospheric molecules and aerosol particles and are reached on the ground the photons are come out from the top of the atmosphere,

(4) Photons are scattered by atmospheric molecules and aerosol particles and are absorbed on the ground.

In the simulation, optical depth of the atmospheric molecules is set at 0.14 while that of the aerosol particles is set at 0.35. Meanwhile, surface reflectance is set at 0.1, 0.15, and 0.2. The percentage ratio of photon pass types is shown in Table 1. The percentage ratios of photon pass types of (1) and (2) are not related to surface reflectance. On the other hands, the percentage ratios of photon pass types of (3) and (4) are depending upon the surface reflectance obviously. Therefore, average ratio of the photons which reaches on the ground depends on surface reflectance as the result.

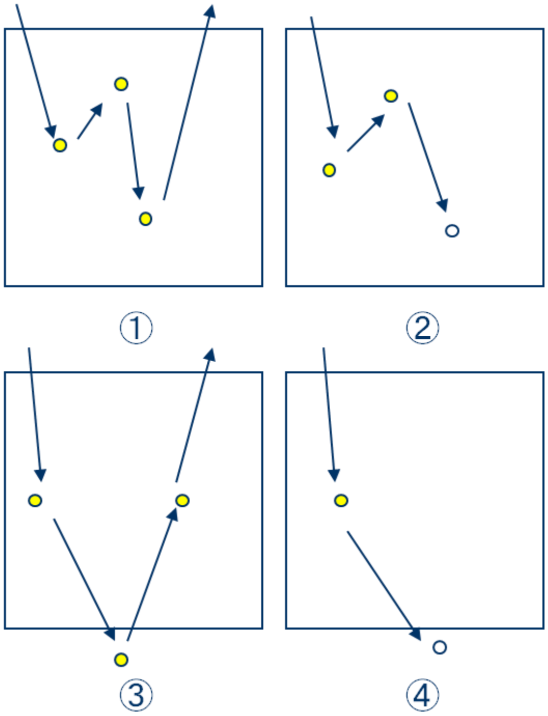


Figure 8. Four types of photons' behavior

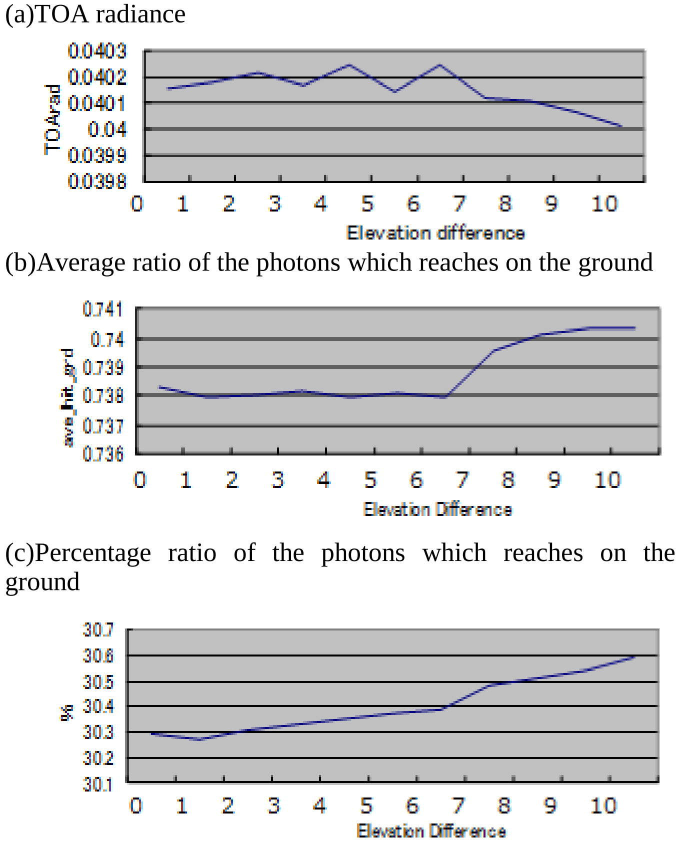
It is obvious that the Top of the Atmosphere: TOA radiance and average ratio of the photons which reaches on the ground depends on surface reflectance and surface elevation differences. Table 1 shows preliminary simulation results. The contribution factors to the TOA radiance are shown in Table 1 as function of surface reflectance. The most dominant factor is No.4 followed by No.1.

The contributions of No.2 and No.3 are almost same and are smaller than those for No.1 and No.4.

TABLE 1. TABLE I. PRELIMINARY SIMULATION RESULTS REFLECTANCE

	Reflectance		
	0.1	0.15	0.2
①	19.30%	19.30%	19.30%
②	8.20%	8.20%	8.20%
③	5.40%	8.22%	11.10%
④	67.1%	64.28%	61.40%
ave_hit_grd	0.739	0.745	0.753
TOArad	0.0401	0.0447	0.0494

Figure 9 (a) shows the TOA radiance as a function of elevation difference while Figure 9 (b) shows the average ratio of the photons which reaches on the ground. Furthermore, Figure 9 (c) shows the percentage ratio of the photons which reaches on the ground. As the results, the percentage ratio of the photons which are reflected on the ground depends on the elevation difference, surface roughness.


(a)TOA radiance

(b)Average ratio of the photons which reaches on the ground

(c)Percentage ratio of the photons which reaches on the ground

Figure 9. TOA radiance and percentage ratio of the photons which reaches on the ground as a function of elevation differences ranges from 0 to 10

Even if the average elevation is same, TOA radiance decreases in accordance with elevation difference due to the fact that photons are reflected several times so that TOA radiance decreases in accordance increasing of surface roughness. Simulation results of the number of photons which are reflected in the atmosphere and on the ground are shown in Table 2.

TABLE 2. TABLE II. SIMULATION RESULTS OF THE NUMBER OF PHOTONS WHICH ARE REFLECTED IN THE ATMOSPHERE AND ON THE GROUND

	Atmosphere		Surface
	Molecule	Aerosol	
Reflectance	1	0.9318	0.1～0.2
Maximum No. of reflected photons	24		4
Average No. of reflected photons	1.9		0.75

There are absorption and scattering due to aerosol particles. Therefore, reflectance of aerosol is not 1.0. 6.72% of photons are absorbed by aerosol particles.

Even though the surface reflectance on the ground is 0.1 to 0.2, 75% of photons are reflected on the ground surface in average.

B. *Simulation Study with a Variety of Parameters*

The following 6 parameters are taken into account in the simulation study,

Standard deviation of surface elevation: S=0.25-1.0, (0.5)

Surface reflectance: ref=0.3-0.7, (0.3)

Optical depth of the atmospheric molecule: tau_ray=0.1-0.35, (0.35)

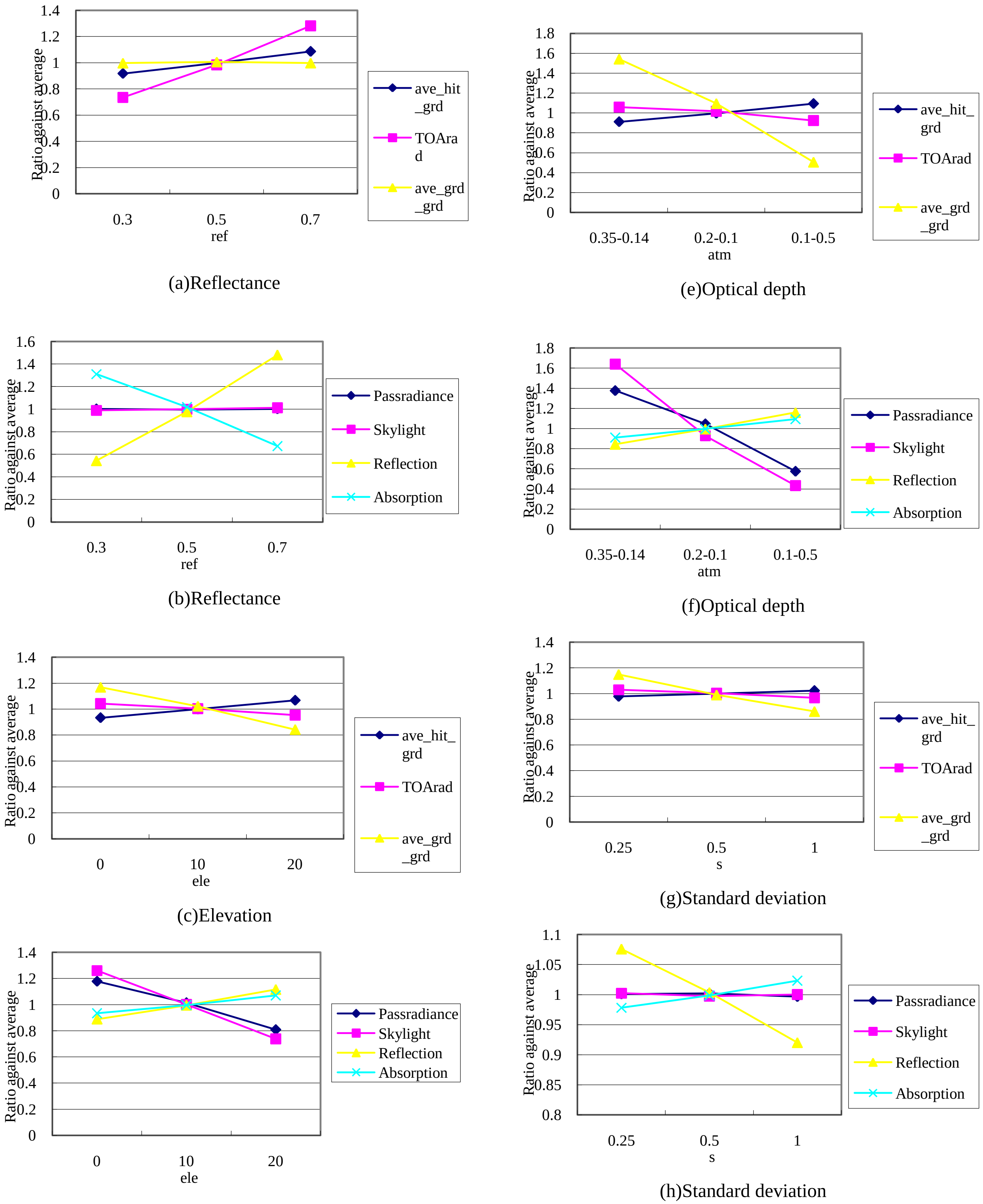
Aerosol optical depth: tau_aero=0.1-0.5, (0.14)

Mean of elevation: ave_Ele=0-20, (10)

Surface slope: 0-30 degree, (0 degree)

where the number in the bracket denotes default values. In the simulation study, average number of photons which hit on the ground, TOA radiance, and average number of photons which reflected on the ground and scattered in the atmosphere then reflected on the ground again are major concerns. Furthermore, pass-radiance, sky-light, reflected radiance on the ground, absorbed radiance in the ground is also concern. Ratio against average of concerned parameters is calculated. The simulation results are shown in Figure 10.

The parameters are r, (1) average number of photons which hit on the ground, (2) TOA radiance, (3) average number of photons which reflected on the ground and scattered in the atmosphere then reflected on the ground again, (4) pass-radiance, (5) sky-light, (6) reflected radiance on the ground, (7) absorbed radiance in the ground.



(a)Reflectance

(e)Optical depth

(b)Reflectance

(f)Optical depth

(c)Elevation

(g)Standard deviation

(d)Elevation

(h)Standard deviation

Figure 10. Ratio against average of concerned parameter

Meanwhile, surface slope effect is evaluated with two adjacent slopes (0 versus 0, 15 versus 0, 30 versus 0, 15 versus 15, 30 versus 15 and 30 versus 30 in unit of degree). Table 3 and 4 shows the results.

TABLE 3. SLOPE EFFECT ON AVERAGE NUMBER OF PHOTONS WHICH HIT ON THE GROUND, TOA RADIANCE, AND AVERAGE NUMBER OF PHOTONS WHICH REFLECTED ON THE GROUND AND SCATTERED IN THE ATMOSPHERE THEN REFLECTED ON THE GROUND AGAIN

	0-0	15-0	30-0	15-15	30-15	30-30
ave_hit_grd	0.8109	0.8114	0.8253	0.8111	0.8204	0.8212
TOArad	0.0548	0.0547	0.0534	0.0546	0.0540	0.0537
ave_grd_grd	1.9744	1.9618	1.8656	1.9605	1.8892	1.8641

TABLE 4. SLOPE EFFECT ON PASS-RADIANCE, SKY-LIGHT, REFLECTED RADIANCE ON THE GROUND, ABSORBED RADIANCE IN THE GROUND

	0-0	15-0	30-0	15-15	30-15	30-30
Passradiance	19230	19222	18544	19230	18913	18865
Skylight	8289	8162	8052	8280	8115	8114
Reflection	15849	15793	15552	15685	15564	15427
Absorption	56632	56823	57852	56805	57408	57594

It is found that multiple reflections on the ground increases in accordance with decreasing the angle between two slopes (absorption in the ground). Accordingly, reflection on the ground decreases. Therefore, scattering in the atmosphere is getting small results in decreasing of the pass-radiance and skylight as shown in Table 4. From the same reason, average number of photons which hit on the ground increases in accordance with two slopes angle is decreased. TOA radiance, and average number of photons which reflected on the ground and scattered in the atmosphere then reflected on the ground again are decreased in accordance with decreasing two slopes angle as shown in Table 3.

When photons reaches at the simulation cell of (1,1), then the photons reflected from the surface as shown in Figure 11. Figure 11 shows the number of photons reflected on the ground. In the simulation, Lambertian surface (iso-tropic reflectance characteristics) is assumed for the ground surface.

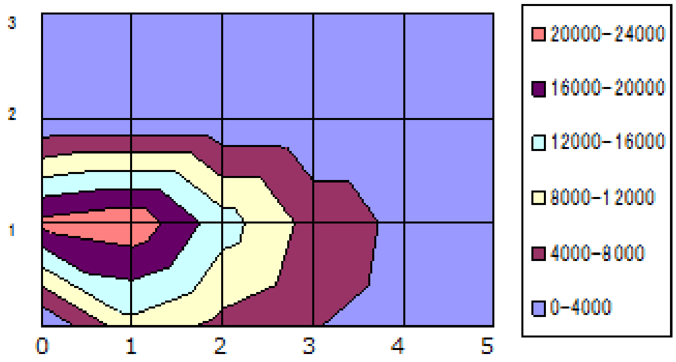
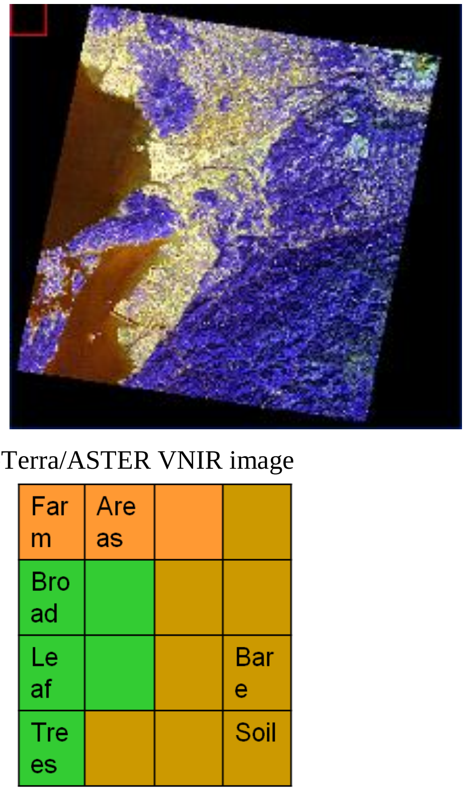


Figure 11. Number of photons reflected on the ground.

C. *Experiemntal Study*

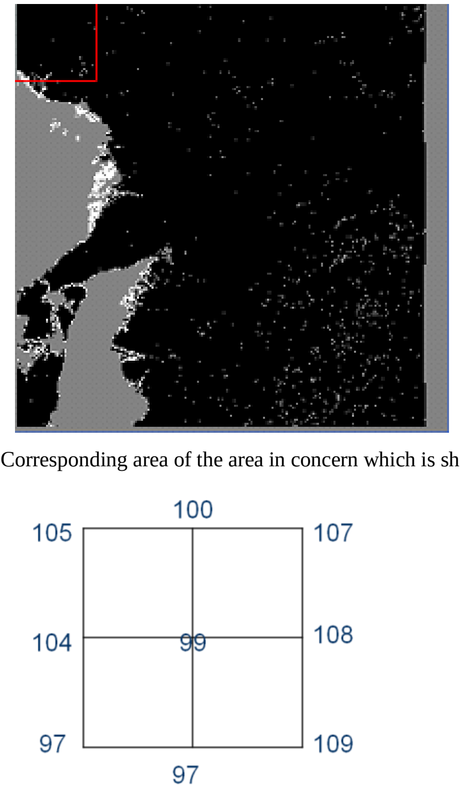
Using Visible to Near Infrared Radiometer: VNIR imagery data of Advanced Spaceborne Thermal Emission and Reflection: ASTER (onboard Terra satellite) [9] Level 3A product (ortho-photo products) of Bands 1, 2, 3N (IFOV of 15m×15m), land cover map (4 by 4 pixels) is created. Also the reflectance of the pixels in concern is estimated using the land cover map as shown in Figure 12.



(a)Terra/ASTER VNIR image

(b)Example of land cover map

Figure 12. Examples of Terra/ASTER/VNIR image and estimated land cover map.



(a)Corresponding area of the area in concern which is shown in Figure 12

(b)Estimated digital elevation level for the four corners of 2 by 2 pixels in concern

Figure 13. Estimated elevations in concern

Utilizing ASTER data product of Level 4A of Digital Elevation Model: DEM (30m×30m), elevations of the pixels in concern are estimated. Thus reflectance of the ground surface can be calculated with 15 m of spatial resolution together with 30 m by 30 m of spatial resolution digital elevations.

Furthermore, 2 by 2 pixels of missing model can be estimated through land cover maps which are created with ASTER/VNIR imagery data based on Maximum Likelihood classification method [10],[11]. Red squares in Figure 12 and 13 are corresponding. Digital elevation at the four corners of 2 by 2 pixels can be calculated using DEM which is derived from Level 4 product of ASTER/VNIR as shown in Figure 13.

Thus the parameters for non-linear mixture model of mixed pixels are determined using MCRT. Then TOA radiance can be estimated precisely taking into account the nonlinearity of the mixels in the TOA radiance calculations with radiative transfer software codes.

IV. Conclusion

Nonlinear mixing model of mixed pixels in remote sensing satellite images taking into account landscape is proposed. Most of linear mixing models of mixed pixels do not work so well because the mixed pixels consist of several ground cover targets in a nonlinear basis essentially.

In particular, mixing model should be nonlinear because reflected photons from a ground cover target are scattered with atmospheric continuants and then reflected by the other or same ground cover targets. Therefore, mixing model has to be nonlinear. Monte Carlo Ray Tracing based nonlinear mixing model is proposed and simulated. Simulation results show a validity of the proposed nonlinear mixed pixel model.

Acknowledgment

The author would like to thank Mrs. Yui Nishimura for her great effort to conduct simulation studies on nonlinear mixture model of mixed pixels of remote sensing satellite images

References

[1] Masao Matsumoto, Hiroki Fujiku, Kiyoshi Tsuchiya, Kohei Arai, Category decomposition in the maximum likelihood classification, Journal of Japan Society of Phtogrammetro and Remote Sensing, 30, 2, 25-34, 1991.

[2] Masao Moriyama, Yasunori Terayama, Kohei Arai, Clafficication method based on the mixing ratio by means of category decomposition, Journal of Remote Sensing Society of Japan, 13, 3, 23-32, 1993.

[3] Kohei Arai and H.Chen, Unmixing method for hyperspectral data based on subspace method with learning process, Techninical Notes of the Science and Engineering Faculty of Saga University,, 35, 1, 41-46, 2006.

[4] Kohei Arai and Y.Terayama, Label Relaxation Using a Linear Mixture Model, International Journal of Remote Sensing, 13, 16, 3217-3227, 1992.

[5] Kohei Arai, Yasunori Terayama, Yoko Ueda, Masao Moriyama, Cloud coverage ratio estimations within a pixel by means of category decomposition, Journal of Japan Society of Phtogrammetro and Remote Sensing, 31, 5, 4-10, 1992.

[6] Kohei Arai, Non-linear mixture model of mixed pixels in remote sensing satellite images based on Monte Carlo simulation, Advances in Space Research, 41, 11, 1715-1723, 2008.

[7] Kohei Arai, Kakei Chen, Category decomposition of hyper spectral data analysis based on sub-space method with learning processes, Journal of Japan Society of Phtogrammetro and Remote Sensing, 45, 5, 23-31, 2006.

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

[9] Ramachandran, Justice, Abrams(Edt.),Kohei Arai et al., Land Remote Sensing and Global Environmental Changes, Part-II, Sec.5: ASTER VNIR and SWIR Radiometric Calibration and Atmospheric Correction, 83-116, Springer 2010.

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

[11] 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.

Passive Clustering for Efficient Energy Conservation in Wireless Sensor Network

Abderrahim MAIZATE
STIC Laboratory
Chouaib Doukkali University, B.P.:20
Eljadida, Morocco

Najib EL KAMOUN
STIC Laboratory
Chouaib Doukkali University, B.P.:20
Eljadida, Morocco

***Abstract*—A wireless sensor network is a set of miniature nodes that consume little energy and route information to a base station. It will enable reliable monitoring of a wide variety of phenomena for civilian, military and medical applications. Almost any sensor network application requires some form of self-organisation to route information. Recent years many protocols for network self-organization and management have been proposed and being implemented. Hierarchical clustering algorithms are very important in increasing the network's life time. The most important point in this algorithm is cluster head selection and cluster formation because a good clustering guarantees reliability, energy efficiency and load balancing in the network. In this paper, we will use the principles of passive clustering to propose a new mechanism for selecting clusterheads. This mechanism allows the election of an alternate for each cluster head and a dynamic balancing of the role of clusterhead to the alternate when leaving or failure. Thus, it provides several advantages network reliability, stability of clusters and reduces energy consumption among the sensor nodes. Comparison with the existing schemes such as Passive Clustering and GRIDS (Geographically Repulsive Insomnious Distributed Sensors) reveals that the mechanism for selecting an alternate for clusterhead nodes, which is the most important factor influencing the clustering performance, can significantly improves the network lifetime.**

Keywords—wireless sensor networks; self-organization; Clustering passive; Clustering; network lifetime; energy efficiency; Fault tolerance; Residual Energy.

I. INTRODUCTION

The wireless sensor networks (WSN) is currently generating a growing interest among researchers. A WSN can be generally described as a network of nodes that cooperatively sense and may control the environment enabling interaction between persons or computers and the surrounding environment [1]. Today, due to recent advances in wireless technologies, new products operating wireless sensor networks are used to retrieve data from these applications. Examples include environmental monitoring, smart homes and offices, surveillance, intelligent transportation systems, and many others (Fig 1).

The use of wireless sensor networks (WSNs) does not require a specific infrastructure. But it poses a problem of scalability, energy conservation and connectivity over time [1], [2], [3]. Also wireless sensor networks can be exposed to highly dynamic and mobile environments, and therefore they must be dynamic and mobile environments, and therefore they must be fault tolerant nodes. Algorithms for wireless sensor networks must be distributed to avoid single point's failure, and self-organization for scalable deployment.

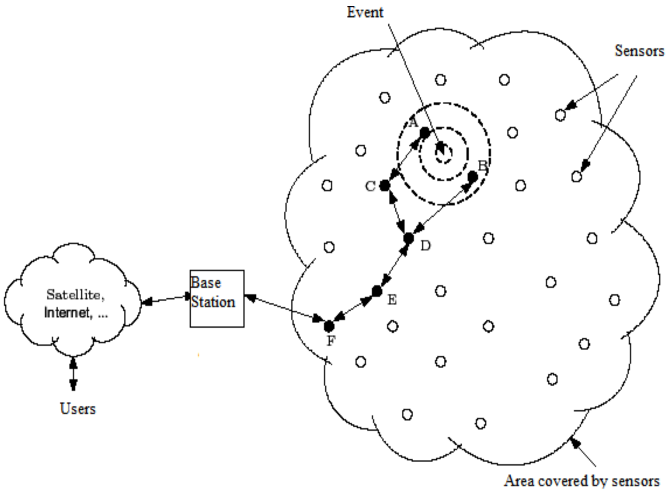


Fig. 1. Traditional scheme of a wireless sensor network.

Self-organization can be defined as the emergence of a global behavior from local interaction [4]. Wireless sensor networks are bandwidth and energy constrained. Self-organization algorithms that minimize the number of message transmissions (and receptions) are preferable. The challenge is to organize dynamic and spontaneous nodes to form a network and prolonging the life-time, while satisfying the constraints of service quality. It is therefore necessary for self-organizing algorithms proposed to limit control packet exchanges for a minimum expenditure of energy and preserve the structure of self-organization for better stability and reliability the network.

How to design an energy efficient protocol to lessen the battery consumption and prolong the network lifetime becomes a critical issue. The passive clustering structure is a way to reduce the energy consumption and can used to perform data aggregation, the process involves grouping nodes into clusters and elect one cluster head (CH) for each cluster to transmit the collected data to central base station through other CHs. Obviously, we can save a great amount of unnecessary transmission by such clustering and data aggregation mechanisms.

A number of recent clustering approach provide techniques to balance energy consumption of network nodes, these techniques are essentially turning the role of CH randomly but it is not fault tolerant nodes especially nodes clusterheads.

The algorithm presented in this paper considers nodes Clusterhead. In this algorithm, we rely on passive clustering to create a redundant role for clusterhead. The node that replaces the cluster head will be the state clusterhead alternate. The clusterhead and the alternate shall periodically to transmit information and verify the presence of each. The proposed algorithm is designed for generating a reliable, scalable and efficient topology, reducing consumption of scarce resources during query dissemination and, thereby extend network lifetime.

The rest of the paper is organized as follows. In Section II, the term self-organization in wireless sensor network is discussed as well as related concepts. We present a description and analysis of the proposed algorithm in Section III, followed by simulation results in Section IV. The last section concludes the paper.

II. Related Work

Clustering is an important research topic in the areas of wireless sensor network (WSN) because clustering improves the performance of many systems. In WSN, clustering can be used to improve the network performance through quality of service metrics such as throughput and delay, in the presence of both mobility and a large number of mobile nodes with minimal resources. We have investigated a number of prior works which consider the wireless sensor network self-organization with clustering mechanisms, a large variety of approaches have been presented by various researchers [5,6, 7,8,9,10,11,12,13,14,14].

Passive clustering [15] can be described as on demand cluster formation protocol that does not use dedicated protocol-specific control packets or signals. The formation of cluster is dynamic and initiated by the first data message to be flooded. Which in turn reduces the duration of the initial set-up period, and the benefits of the reduction of the forwarding set can be felt by calculating the total energy consumed because the main function of the clusters is to optimize the exchange of flooded messages.

In passive clustering each node operates the MAC sender address carried by the received packets to collect neighbor information, and can construct clusters even without collecting the complete neighbor list. Instead of using protocol specific signals or packets, passive clustering reserves two bits for the following four states of a mobile node: :

(1) Initial, (2) Cluster head, (3) Gateway and (4) Ordinary. At the beginning, every sensor node is in the INITIAL state until it receives a packet. If the sender's packet is not CLUSTERHEAD, this sensor node switches to CLUSTERHEAD-READY. This node will become a CLUSTERHEAD if it successfully transmits a packet before receiving any packets from others. If the sensor node receives a packet from a CLUSTERHEAD it changes state to ORDINADRY. Any sensor node that hears more then one CLUSTERHEAD becomes GATEWAY.

Passive clustering has several mechanisms for the cluster formation such as: Gateway Selection Heuristic and First Declaration Wins rule. The Gateway Selection Heuristic provides a procedure to elect the minimal number of gateways required to maintain the connectivity between clusterheads. With the First Declaration Wins rule, a node that first transmits a data message will be a clusterhead of the rest of nodes in its clustered area.

Passive clustering maintains clusters using implicit timeout. A node assumes that some nodes are out of clustered area if they have not sent any data longer than timeout duration. With reasonable offered load, a node can catch dynamic topology changes.

In [16], an interesting technique has been proposed "Energy Conserving Passive Clustering (ECPC) algorithm" which takes account of both residual energy and distance for becoming cluster head and gateway and also eliminate the problem of idle listening through periodic sleep and awake among the cluster members. This algorithm outperforms Directed Diffusion (DD) [17] and Passive Clustering Directed Diffusion (PCDD) [18] in terms of energy dissipation and network lifetime. It also generates much less gateway nodes than PCDD algorithm. The selection of the clusterhead is based on the higher residual energy with in the 1-hop neighbours and the distance to form better clusters. The gateway selection procedure prioritizes the residual energy of the node which wants declare itself as a gateway and average distance of clusterhead nodes with in -hop.

Some of the well known clustering protocols are the Low-Energy Adaptive Clustering Hierarchy (LEACH) [19] and the Hybrid-Energy-Efficient Distributed (HEED) [20]. Both of which are self-organizing, and distributed protocols. LEACH achieves energy saving in three ways: randomized rotation of cluster head, sleep mode and data aggregation. CHs are randomly selected. The decision of CHs is simply based on the suggested percentage of them for the network and the number of times the node has been a CH therefore two ch can be selected in close area, thus the system efficiency may be decreased. There are several variant of this algorithm such as: LEACH-B, LEACH-C, LEACH-E and M-LEACH.

HEED improves network lifetime over LEACH by distributing energy consumption. HEED focuses on choose appropriate CHs by using residual energy as the primary clustering parameter to select a number of tentative CHs. Those tentative CHs inform their neighbours of their intentions to become CHs. These advertisement messages include a secondary cost measure that is a function of neighbor proximity or node degree. This secondary cost is used to help the regular nodes in choosing the best clusterhead to join, and to avoid elected CHs being within the same cluster area of each other. If a CH is far from the sink, it tries to send the aggregate data to another CH instead of sending to the sink directly.

GRIDS [21] is an energy-aware cluster formation protocol which increase network lifetime by using an efficient selection mechanism of critical (or not) nodes. This mechanism allows balanced energy consumption among the sensor nodes without requiring additional overheads including additional signaling, time synchronization and global information. GRIDS is based on an energy model which delivers node’s remaining energy level in real time. This information is piggybacked in the nodes packet header. Each sensor determines being insomnious or not based on its remaining energy and the number of neighbouring insomnious nodes and their energy level. An efficient flooding during each wake up period determines insomnious nodes in the network.

GRIDS inherit PC for constructing and maintaining clusters. The most important advantage of GRIDS compared to PC is that a set of nodes in a cluster with higher energy levels have higher probability to become critical nodes, i.e., CH or GW. In Passive Clustering, CHs keep their cluster status until there is a CH collision, i.e. the hop distance between two CHs becomes 1, and one of them resigns from CH. In GRIDS, an energy abundant node can challenge CH and usurps the role. Even if there is a CH declaration, nodes can challenge when their energy levels are higher than the one of CH. These nodes keep their cluster status even if they receive packets from the current CH.

Clustering stability, fast convergence time, the consumption of energy and mobility of nodes are important properties required of clustering algorithms. To improve the stability and reliability of clustering and reduce consumption of energy, we developed a new rule of electing alternates of clusterheads that represent critical nodes for passive clustering.

III. Proposal- Passive Clustering for Efficient Energy Conservation in Wireless Sensor Network

In this section, we present the details of new algorithm. The advantage is that improves reliability of the network by selecting an alternate for clusterhead, uses balanced energy consumption among network nodes and keeps longer the structure of clusters.

As a result, the network stability and reliability are preserved, the transmission delay is decreased and the life time of the network is significantly increased.

A. PCEEC mechanism

PCEEC (Passive Clustering for Efficient Energy Conservation in Wireless Sensor Network) defines a protocol for cluster formation and election of alternates of the clusterheads based on the following principles:

1) There are six possible states: dead, initial, ordinary, clusterhead_ready, custerhead, gateway and clusterhead-alternate

2) Initially or when there are no networks activities for a long time, all nodes are in the 'initial' state. This state does not change as long as a node does not receive a packet from another node.

3) When a node receives a packet and if the state of a sender is not ClusterHead, the receiver's state.

4) switches to ClusterHead_ready, otherwise the node switches to state ordinary or gateway.

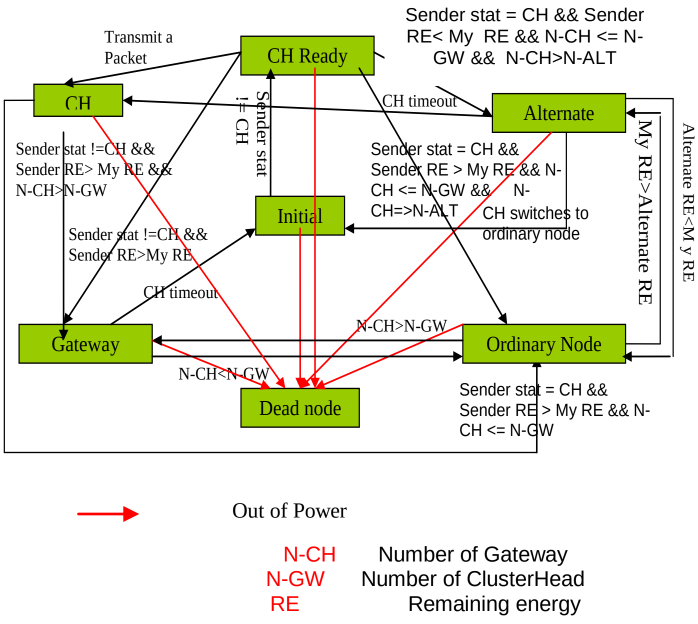

Transmit a Packet
CH Ready
Sender stat = CH && Sender RE< My RE && N-CH <= N-GW && N-CH>N-ALT
CH
CH timeout
Alternate
Sender stat != CH
Sender stat !=CH && Sender RE> My RE && N-CH>N-GW
Sender stat = CH && Sender RE > My RE && N-CH <= N-GW && N-CH=>N-ALT
CH switches to ordinary node
My RE>Alternate RE
Alternate RE<M y RE
Initial
Sender stat !=CH && Sender RE>My RE
CH timeout
N-CH>N-GW
Gateway
Ordinary Node
N-CH<N-GW
Dead node
Sender stat = CH && Sender RE > My RE && N-CH <= N-GW
Out of Power
N-CH Number of Gateway
N-GW Number of ClusterHead
RE Remaining energy

Fig. 2. State diagramme of PCEEC

5) The node ClusterHead_ready switches to state gateway when the number of ClusterHeads is greater or equal to the number of Gateways. Otherwise, the node becomes an Ordinary Node or an alternate node.

6) The node ClusterHead_ready switches to state alternate when the number of ClusterHeads is greater or equal to the number of Gateways and the number of ClusterHeads is greater to the number of alternates. Otherwise, the node becomes an Ordinary Node.

7) The clusterHead chooses the node with the highest level of energy as an alternate on failure the previous. The clusterHead checks periodically the presence of his alternate. In the case of the leaving or failure of the alternate, the cluster head rerun the selection process of a new alternate.

8) Similarly, if the alternate discovers the leaving or failure of cluterhead it switches to state ClusterHead and launch the procedure to select an alternate (see Figure 2).

9) An ordinary node switches to alternate if its energy is higher. The alternate node switches to state ordinary.

PCEEC uses the same principles as PC for the construction and maintenance of clusters in wireless sensor networks. It also inherits the characteristics of the algorithm GRIDS by giving nodes with the highest level of energy to become a critical node, i.e., ClusterHead, Alternate or GateWay.

B. Operational Description

In each cluster, we will have a cluster head that centralizes information and an alternate to replace him on failure. Thus the structure of the cluster will be further preserved.

TABLE I. A PSEUDO-CODE THAT SHOWS THE OPERATION DETAILS OF PCEEC.

```
                         INPUT
              NBALT: Number of alternates.
                         BEGIN
/* RE=Remaining energy*/
/* initially, all nodes are in the initial state */
   Node.state = Initial;
If node.energy=0
   Node.state=Dead.
If  Node .state = Initial {
     If Sender.state ! = ClusterHead
                    Node .state = ClusterHead Ready

      If  (Sender.state = ClusterHead && Sender remaining
      energy > My  remaining energy)
         Node .state = Gateway
    }
Else
     If  Node .state = ClusterHead Ready {
           If  (Node transmits a packet)
               Node .state = ClusterHead
         If (Sender.state = ClusterHead  && Sender
             remaining energy > My  remaining energy &&
             NBR(ClusterHead) <= NBR(Gateway) &&
             NBR(CH) > NBR(ALT))
                Node .state = Ordinary Node
           If (Sender.state = ClusterHead && Sender
             remaining energy < My  remaining energy &&
             NBR(ClusterHead) <= NBR(Gateway) &&
             NBR(Alternate) <=NBR(CH))
                Node .state = Alternate
            If  (Sender.state = ClusterHead && Sender
             remaining energy > My  remaining energy &&
             NBR(ClusterHead) > NBR(Gateway))
                Node .state = Gateway
     }
    Else
          If  Node .state = ClusterHead {
             If (Sender.state = ClusterHead && Sender
                 remaining energy > My  remaining energy
                 && NBR(ClusterHead) <= NBR(Gateway))
                      Node .state = Ordinary Node
                       ClusterHead-alternate.state= Ordinary
                       Node
             If  (Sender.state = ClusterHead && Sender
                 remaining energy > My  remaining energy
                 && NBR(ClusterHead) <= NBR(Gateway))
                      Node .state = Gateway
              If  Clusterhead-alternate  TimeOut
                   Launch the procedure the selection of
                   alternate
           }
          Else
             If  Node .state = Ordinary Node  {
                 If CH TimeOut
                         Node .state = Initial
          If  NBR(ClusterHead) > #(Gateway)
                         Node .state = Gateway
         If  My RE > Alternate RE
                        Node .state = Alternate

                 }
                 Else
        If  Node .state = Gateway {
                If CH TimeOut
                        Node .state = Initial
                      If  NBR( ClusterHead) <= NBR( Gateway)
                           Node .state = Ordinary Node
                          }
                 Else
          If Node .state = ClusterHead-Alternate  {
              If  CH TimeOut
                  Node .state = ClusterHead
              If My RE< RE of an ordinary node
                   Node .state = Ordinary

}
```

IV. SIMULATIONS

In this section, we present comparison between proposed algorithms and two most important clustering protocols, PC and GRIDS. This comparison is evaluated using the simulator GLOMOSIM [22], which is a scalable simulation environment for wireless networks based on the Parsec language [23]. We begin first by specifying the metrics that we considered interesting to evaluate this algorithm and results obtained.

A. *Metric*

To evaluate the performance of PCEEC protocol, we will use the following parameters:

- Lifetime of the network: duration until the death of the last node in the network.
- Delivery rate: it characterizes the routing performance and is based on network conditions. It is the fraction effective throughput / maximum flow.
- Energy consumption: Amount of energy needed to sustain the network during its lifetime and data collection.
- Dead Nodes: Represents the percentage of dead nodes over time in the wireless sensor network.

B. *Simulation scenarios*

The simulation parameters used are as follows:

- the roaming space is 500X500 m square,
- The radio propagation of each node reaches up to 250 meters
- The channel capacity is 2 Mbits/second.
- The battery capacity is equal to 500 mW

- Simulations use a variable number of nodes; distributed randomly in the roaming area;
- The random-way point model is used for node mobility
- The traffic model used is constant bit rate (CBR). Packet length is 566 bytes.
- Each node sends 100 packets with inter-arrival time of 0.2 second.
- Number of nodes: 300.
- AODV [24] is chosen as the routing protocol;

We use four metrics for analyze and compare the simulation results: network lifetime, consumed energy, Percentage of dead nodes and delivery ratio at base station.

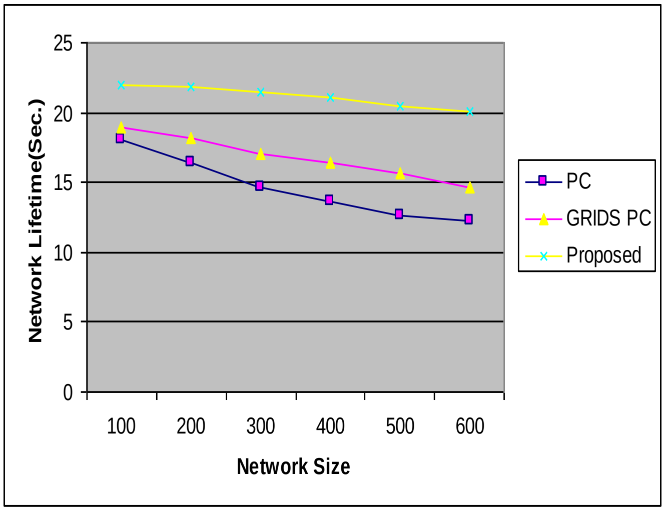


Fig. 3. Network lifetime of PCEEC compared to PC and GRIDS PC.

Figure 3 plots the network size and the network lifetime for the three protocols. We measure the network lifetime when the number of sensors varies between 50 and 600. The energy consumption model is linear. The results show that PCEEC achieves better performances compared to the two others. The PCEEC protocol improves the network lifetime performance, and the gain in lifetime increases with the size of the network. Thus, we conclude that this protocol is more suitable for large scale networks.

Figure 4 shows that the proposed algorithm consumes less energy than the passive clustering and GRIDS PC as can be seen. The PCEEC protocol saves 20% of the total energy consumed, because our algorithm preserves better cluster structure by preventing reinitialization in case of the leaving or failure of a clusterhead.

Figure 5 indicates the energy consumption, during the phase of reclustering, generated by PC, GRIDS PC and proposed PCEEC algorithm in case the failure the node clusterhead. It is obviously that the proposed algorithm reduces significantly the consumed energy with a percentage very close to 50%. This gain is due to the preservation of the cluster structure.

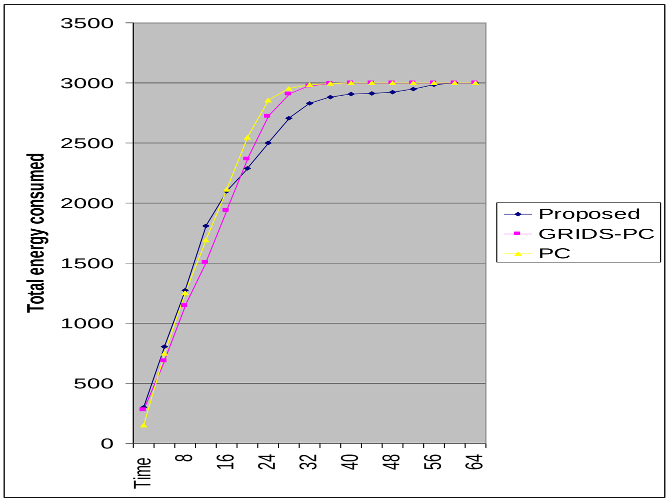


Fig. 4. Total energy consumed by PCEEC compared to PC and GRIDS PC.

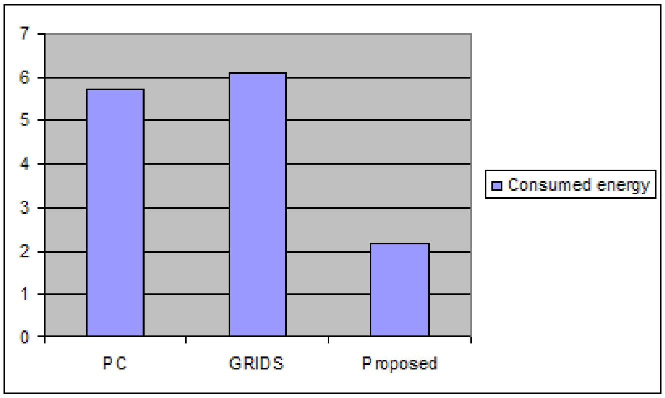


Fig. 5. Consumed energy by the reclustering after the failure of the CH in a simulation of 300 nodes.

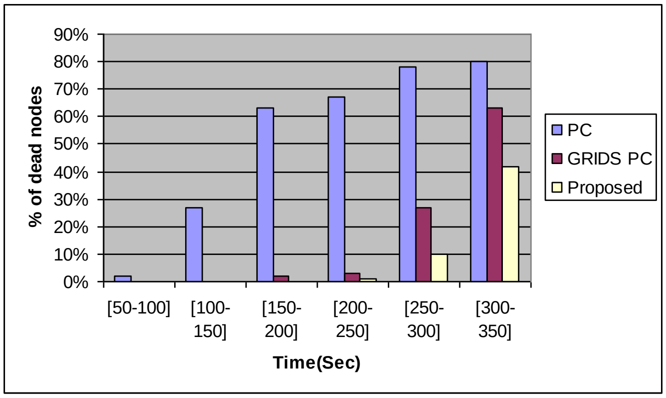


Fig. 6. Percentage of dead nodes in a simulation of 300 nodes.

Figure 6 shows dead nodes ratio as a function of time. PCEEC outperforms and achieves better results in optimizing the energy consumption compared to passive clustering and GRIDS PC.

Similarly, Fig. 7 shows that also the Delivery ratio is much better with PCEEC, because PCEEC decreases the number of dead nodes and retains more the cluster structure.

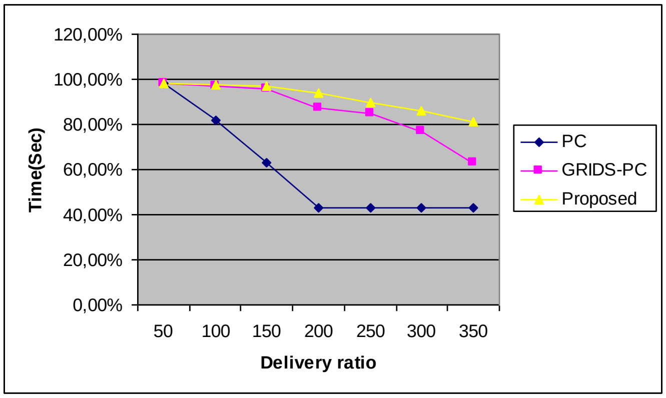


Fig. 7. Delivery ratio in a simulation of 300 nodes.

Thus, the simulation results show that the Passive Clustering for Efficient Energy Conservation in Wireless Sensor Network scheme not only provides an efficient forwarding and balances the energy consumption but also improves network performance.

V. Conclusion and Future Work

We introduce a passive clustering mechanism for electing clusterheads and clusterheads alternate in wireless sensor networks. The selection of cluster heads and alternate is performed according to the remaining energy of sensor nodes. The sensor nodes with the highest energy in the clusters can be a cluste headsand alternates at different cycles of time. Thus, the role of cluster heads and alternate can be switched dynamically. Simulation results show the effectiveness of the approach in reducing the amount of energy consumed by the network in comparison with two well-known protocols, passive clustering and GRIDS PC.

In the future, it is planned to provide enhancements to the proposed algorithm to make decisions using distance during the selection of clusterheads and alternates to improve the performance further.

References

[1] Verdone, R.; Dardari, D.; Mazzini, G.; Conti, A. Wireless Sensor and Actuator Networks; Elsevier: London, UK, 2008.

[2] Akyildiz I. F., Y. Sankarasubramaniam W. Su. et Cayirici E. Wireless sensor networks : a survey. Computer Networks (Elsevier), vol. 38(4), pp. 393–422, March. 2002.

[3] Yick J., Mukherjee B. et Ghosal D. Wireless sensor network survey. Computer Networks (Elsevier), vol. 52, n° 12, pp. 2292–2330, Aug. 2008.

[4] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energyefficient communication protocol for wireless microsensor networks,” in Proc. of the 33rd Annual Hawaii International Conference on System Sciences (HICSS’00), Hawaii, USA, Jan. 2000, pp. 3005–3014.

[5] Houda Zeghilet; Moufida Maimour; Nadjib Badache; Francis Lepage. On the Use of Passive Clustering in Wireless Video Sensor Networks Int. J. of Sensor Networks, 2012 Vol.11, No.2, pp.67-80.

[6] Ketki Ram Bhakare,R. K. Krishna,Samiksha Bhakare, “An Energy-efficient Grid based Clustering Topology for a Wireless Sensor Network,” International Journal of Computer Applications. Volume 39– No.14, February 2012.

[7] Rajnish Kansal, “Enhanced Uniform Distributed Clustering Algorithm (UDCA) In Wireless Sensor Network,” International Journal of Emerging Technology and Advanced Engineering, Volume 2, Issue 6, June 2012.

[8] Zhu Yonga, Qing Peia, “A Energy-Efficient Clustering Routing Algorithm Based on Distance and Residual Energy for Wireless Sensor Networks,” 2012 International Workshop on Information and Electronics Engineering (IWIEE).

[9] M. H. Tolou and J. Chitizadeh, “Lifetime prolonging of wireless sensor networks via a recursive clustering algorithm,” in Proc. of the third IEEE International Conference in Central Asia on internet the Next generation of mobile,wireless and optical communications networks (IEEE/IFIP ICI’07), Tashkent, Sep. 2007, pp. 1–6.

[10] C. H. and M. S., “Cluster sizing and head selection for efficient data aggregation and routing in sensor networks,” in Proc. of IEEE Wireless Communications and Networking Conference (IEEE WCNC’06), vol. 4, Las Vegas, NV, USA, 2006, pp. 2318–2323.

[11] Y. Zhou, M. Hart, S. Vadgama, and A. Rouz, “A hierarchical clustering method in wireless ad hoc sensor networks,” in Proc. of the IEEE International Conference on Communications (IEEE ICC’07), Glasgow, Scotland, Jun. 2007, pp. 3503–3509.

[12] J. Yu, W. Liu, J. Song, and B. Cao, “Eemr: An energy-efficient multi-hop routing protocol for wireless sensor networks,” in Proc. Of the International Conference on Computer Systems and Application (IEEE/ACS AICCSA’08), Doha, Qatar, Mar. 2008, pp. 291–298.

[13] M. H. Yeo, M. S. Lee, S. J. Lee, and J. S. Yoo, “Data correlation-based clustering in sensor networks,” in Proc. of the International Symposium on Computer Science and its Applications (CSA’08), Hobart, Australia, Oct. 2008, pp. 332–337.

[14] Amir Akhavan Kharazian, K. Jamshidi and M. Khayyambashi, “adaptive clustering in wireless sensor network: considering nodes with lowest energy,“ International Journal of Ad hoc, Sensor & Ubiquitous Computing (IJASUC) Vol.3, No.2, April 2012.

[15] M. Gerla, T.J. Kwon and G. Pei “On Demand Routing in Large Ad Hoc Wireless Networks with Passive Clustering”, Proceedings of IEEE WCNC 2000, Chicago, IL, Sep. 2000.

[16] Md. Mamun-or-Rashid, M. Mahbub Alam and C. Seon Hong, “Energy Conserving Passive Clustering for Efficient Routing in Wireless Sensor Network,”

[17] C. Intanagonwiwat, R. Govindan, and D. Estrin, Directed diffusion: a scalable and robust communication paradigm for sensor networks, Proceedings of ACM MobiCom '00, Boston, MA, (2000) 56-67

[18] Handziski V., Köpke A., Karl H., Frank C. and Drytkiewicz W. “Improving the Energy Efficiency of Directed Diffusion Using Passive Clustering”, in Proc of the 1st European Workshop on Wireless Sensor Networks (EWSN), LNCS 2920, Berlin, Germany, (2004) 172-187

[19] W. Heinzelman, A. Chandrakasan and H. Balakrishnan, Energy-Efficient Communication Protocol for Wireless Micro-sensor Networks, Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS '00), (2000)

[20] O. Younis and S. Fahmy, “HEED: A Hybrid, Energy-efficient, Distributed Clustering Approach for Ad Hoc Sensor Networks,” IEEE Transactions on Mobile Computing, vol. 3, no. 4, pp. 366-379, 2004.

[21] El Ghanami, D. Kwon, T.J. ; Hafid A., “ GRIDS: Geographically Repulsive Insomnious Distributed Sensors – An Efficient Node Selection Mechanism Using Passive Clustering,” Networking and Communications, 2008. WIMOB '08. IEEE International Conference on Wireless and Mobile Computing,

[22] M. Gerla, L. Bajai, M. Takai, and R. Ahuja. GloMoSim: A Scalable Network Simulation Environment, Technical Report 990027, University of California at Berkley, 1999.

[23] R. Bagrodia, R. Meyer, M. Takai, Y. Chen, X. Zeng, J. Martin, B. Park, H. Song, Parsec: A parallel simulation environment for complex systems, Computer, Vol. 31(10), October 1998, pp. 77-85.

[24] C. E. Perkins et al., Ad hoc on-demand distance vector (AODV) routing, [Online] Available: http://www.ietf.org/internet-drafts/draft-ietf-manet-aodv-13.TX

Inferring the Human Emotional State of Mind using Assymetric Distrubution

N. Murali Krishna[1]
Asst Professor
Dept of CSE, GITAM University
Vishakhapatnam, INDIA

P.V. Lakshmi [2]
Professor
Dept of IT, GITAM University
Vishakhapatnam, INDIA

Y. Srinivas [3]
Professor
Dept of IT, GITAM University
Vishakhapatnam, INDIA

Abstract— **This present paper highlights a methodology for Emotion Recognition based on Skew Symmetric Gaussian Mixture Model classifier and MFCC-SDC ceptral coefficients as the features for the recognition of various emotions from the generated data-set of emotional voices belonging to students of both genders in GITAM University. For training and testing of the developed methodology, the data collection is carried out from the students of GITAM University of Visakhapatnam campus using acting sequence consisting of five different emotions namely Happy, Sad, Angry, Neutral, Boredom; each uttering one short emotional base sentence. For training the data we have considered fifty speakers from different regions (30 male & 20 female) and one long sentence containing an emotional speech from each speaker. The experimentation is conducted on text dependent speech emotion recognition and results obtained are tabulated by constructing a Confusion Matrix and comparing with existing methodology like Gaussian mixture model.**

Keywords— ***Skew Gaussian Mixture Model; MFCC; SDC; Emotion Recognition; Confusion Matrix***

I. INTRODUCTION

For human communication system, speech acts as the medium from his/her voices. The main advantage of speech is that we can identify the interacting person voice. Emotion is a integral part of the speech which helps to identify the internal feelings of the speaker and in other words, emotion helps to understand the listeners state of mind. In many of the practical situations like BPO, telephonic communication etc it is desirable to understand the emotions of the speaker [1]. The Emotion Recognition is carried out by an acting sequence generated by each speaker varying different emotion back grounds, these speeches are therefore termed as acting sequences.

However, while generating different emotions the same step is to be applied. The emotions of speech narrate the prosody in a speech. The prosody of speech depends upon many characters, which include rules of the language, condition of living, place of living, culture of community etc [2]. Many models have been evolved for the recognition of the emotions based on GMM [3, 4, 5, 6], SVM [7, 8], HMM [9, 10], Truncated GMM [11, 12]. In order to have effective speaker emotion recognition system, the features that help in recognition the emotions are to be identified effectively. Many of the models in Literature use MFCC coefficients for the recognition of speaker emotions.

However, these MFCC coefficients will be useful, if speech is of short duration, for long-term speech Signals Shifted Delta Coefficients (SDC) features are more appropriated since they identify the dynamic behavior of the speaker along with the prosodic features of the emotions. More over the models discussed mainly in the Literature Review are aimed towards the identification of the speaker’s emotional speech, assuming that the emotion speech signals are symmetric. However in reality, the speech signal is asymmetric in nature.

Hence in order to interpret the speech signal effectively, it is advantageous to use asymmetric distributions like Skew Gaussian Distribution, Log Distribution, Gamma Distribution, etc. In this paper we have utilized Skew Gaussian Distribution, since it contains Gaussian distribution as a particular case. The rest of the paper is organized as follows, in Section-2, the feature extraction methodology is presented, section -3, deals with the skew Gaussian mixture model, methodology is presented along with performance evaluation in section -4, the section-5 of the paper deals with results arrived and comparisons.

II. THE FEATURE EXTRACTION

For effective recognition of emotional speech it is important to extract the features effectively. In this paper we have considered the MFCC-SDC features for the extraction of features from the speech database. MFCC are preferred because of its ability to interpret the signal in short duration and SDC is used to extract the features from the long duration speech samples or dynamically changing samples. The processes carried out for the extraction of features are

1. The speech voices are fragmented in to small frames and these frames are given to MFCC.

2. Long term speeches are also considered and segmented the speech samples of windows sizes of 30ms, 60 ms etc and these sequences are given to SDC.

3. Using the combo effect of MFCC-SDC, the features are extracted for different emotions.

4. Classification of the emotion is carried out using Skew Gaussian mixture model.

5. The developed model is compared to that of GMM.

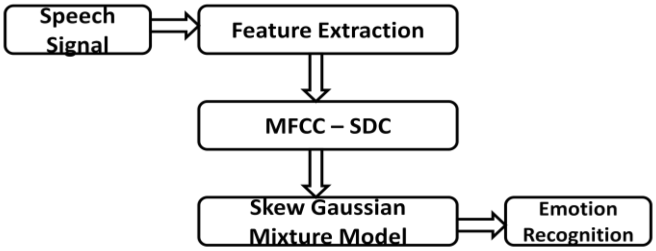


Fig-1 Process of Emotion Recognition Model

The speech signal in the database containing different emotions such as Happy, Sad, Angry, Neutral, Boredom are extracted and are trained using Skew Gaussian mixture model. The feature extraction process involves the generation of emotions samples in .WAV form extracting the amplitude values for each of these emotions signals and these values are given as input to be Skew Gaussian distribution and the probability density function values of the Skew Gaussian distribution are generated. For the testing purpose, test signal is considered, amplitude are values generated and these feature values are to be given as input to the speech signal for classification of the emotion.

III. SKEW GAUSSIAN DISTRIBUTION

For the recognition of speech samples it is essential to understand the behavioral pattern of the speech signal. In many of the cases, the speech signal assumed to be symmetric but speech signal mainly depends upon the prosodic and also depends upon the pitch, energy and other factors associated with each speaker. Since these features are not symmetric always, asymmetric features are to be considered. Hence in this paper we have considered speech signal which caters the speeches which are both symmetric and asymmetric.

The probability density function of Skew Gaussian distribution is given by

$$f(\mathrm{z}) = 2.\emptyset(z).\emptyset(\alpha\, Z); \quad -\alpha < Z < \alpha \tag{1}$$

Where, $\emptyset(\alpha\, z) = \int_{-\alpha}^{\alpha z} \emptyset(t)dt$ (2)

And, $\emptyset(z) = \frac{e^{\frac{1}{2}z^2}}{\sqrt{2\pi}}$ (3)

Let $y = \mu + \sigma z$

$$Z = \frac{y-\mu}{\sigma} \tag{4}$$

Here we are submitting the equations (2),(3), and (4) in equation (1), we have

$$f(z) = \sqrt{\frac{2}{\pi}}\ e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2} \left| \int_{-\alpha}^{\alpha(\frac{y-\mu}{\sigma})} \frac{e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2}}{\sqrt{2\pi}} dt \right| \tag{5}$$

IV. EXPERIMENTATION EVALUATION

To demonstrate the model a database is generated with 50 students (30 male and 20 female) of GITAM University containing different dialects of Andhra Pradesh. The voices are recorded in acting sequences with different emotions Happy, Sad, Angry, Neutral and Boredom. The process of the emotion classification is as follows

A. Phase-1Extract the MFCC-SDC coefficients

B. Phase-2: Train The Data By The Probability Density Function Of Skew Gaussian Distribution Model

C. Phase-3: Consider A Test Signal From Voice Database, Apply The Step-1 & Step-2 And Classify The Emotion

D. Phase-4: The Output Generated Is Depicted In The Form Of Confusion Matrix Shown In Below Table-1 And Table -2

TABLE 1. COMPARISON OF CONFUSION MATRIX FOR IDENTIFY DIFFERENT EMOTION OF MALE

stimulation	Recognition Emotion (%)/*proposed model*					Recognition Emotion (%)/*GMM*				
	Angry	Boredom	Happy	Sadness	Neutral	Angry	Boredom	Happy	Sadness	Neutral
Angry	92	0	0	8	0	80	0	10	10	0
Boredom	5	85	0	10	0	10	70	0	20	0
Happy	0	0	90	0	10	10	10	70	0	10
Sadness	0	10	0	86	4	10	0	10	60	20
Neutral	0	10	0	8	82	0	10	0	20	70

Figure 1. BARCHART-1, REPRESENTING THE RECOGNITION RATES FROM MALE DATABASE

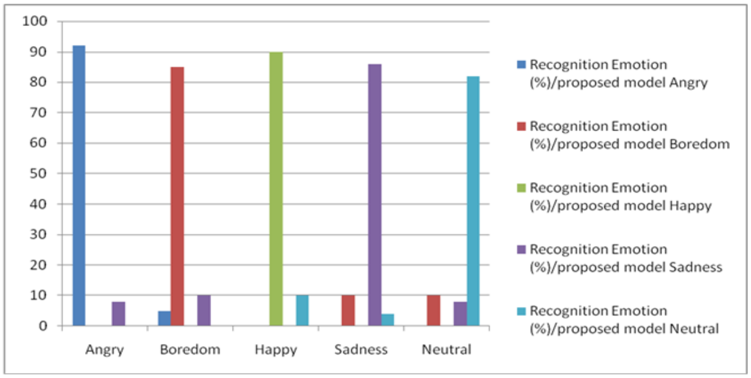
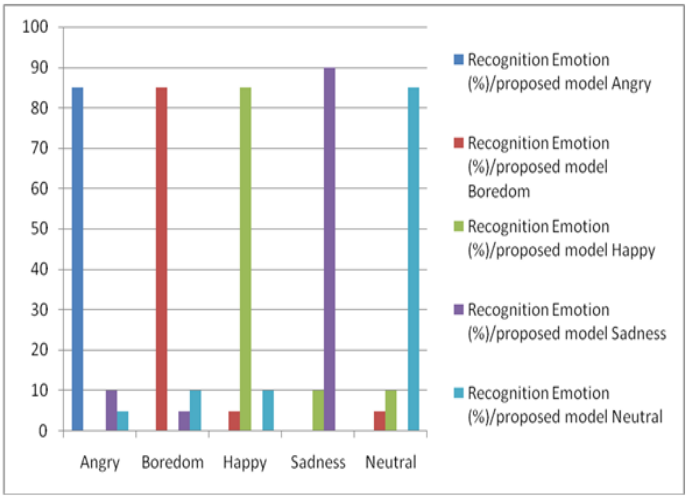


TABLE 2. CONFUSION MATRIX FOR IDENTIFY DIFFERENT EMOTION OF FEMALE

stimulation	Recognition Emotion (%)/*proposed Model*					Recognition Emotion (%)/*GMM*				
	Angry	Boredom	Happy	Sadness	Neutral	Angry	Boredom	Happy	Sadness	Neutral
Angry	85	0	0	10	5	92	0	8	0	0
Boredom	0	85	0	5	10	10	70	20	0	0
Happy	0	5	85	0	10	10	0	82	0	08
Sadness	0	0	10	90	0	10	10	0	70	10
Neutral	0	5	10	0	85	10	0	10	0	80

Figure 2. Barchart-2, Representing The Recognition Rates From Female Databas



The results obtained are compared with that of existing model like Gaussian mixture model and results obtained in the presented in the above Table-1 & Table -2

V. Conclusion

In this paper a novel methodology for Emotion Recognition is done by using Skew Gaussian Mixture Model is developed. These emotions are recorded at 30 ms with five different emotions. The speech database is generated from the acting sequence of one short emotionally based speech sentence comprising of 5 different emotions from50 students (speakers) from different dialects of Andhra Pradesh. The features are extracted and for recognizing, the test speaker's emotion is considered and classified using Skew Gaussian Mixture Model. The results obtained are presented in the confusion matrix for both genders in Table-1 and in Table -2, and Bargraphs-1 &2. From the above Tables and Bar-Graphs, it can be see that the recognition rate is 90%.in case of certain emotion and for the other emotion. The output is compared with that of the existing model based on GMM and from the Table-1 & Table -2, it can be clearly seen that our method outperforms the existing model. The overall emotion rate is above 85%.

References

[1] N. Murali Krishna et al "Emotion Recognition using Dynamic Time Warping Technique for Isolated Words",IJCSI-vol-8,Issue-5,Sept-2011.

[2] björn schuller, et al" speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine - belief network architecture" IEEE, icassp 2004,pp 577-580.

[3] Y.G. Kang et al "A hybrid GMM and codebook mapping method for spectral conversion" in Proc. The 1st International Conference on Affective Computing and Intelligent Interaction, 2005, pp.303-310.

[4] RajeswaraRao.R.,Nagesh et al "Source Feature Based Gender Identification System Using GMM" International Journal on computer science and Engineering,Vol:3(2),2011,pp-586-593.

[5] Xianglin Cheng et al "Speech Emotion Recognition Using Gaussian Mixture Model" International Conference on Computer Application and System Modeling-(2012)

[6] Bihar Kandali Emotion recognition from Assamese speeches using MFCC features and GMM classifier "TENCON, 2008 , Page(s): 1 - 5

[7] Meng-Ju Han"A new information fusion method for SVM-based robotic audio-visual emotion recognition "IEEE International Conference, 2007 , Page(s): 2656 - 2661

[8] Peipei Shen et al "Automatic Speech Emotion Recognition using Support Vector Machine "EMEIT, 2011 , Page(s): 621 – 625

[9] Xia Mao et al "Multi-level Speech Emotion Recognition Based on HMM and ANN", 2009 WRI World Congress,Computer Science and Information Engineering, pp.225-229, March2009.

[10] K. Choi and J.N. Hwang, "Baum-Welch HMM inversion for reliable audio-to-visual conversion", in Proc. IEEE International Workshop Multimedia Signal Processing, 1999, pp.175-180

[11] N. Murali Krishna et al "An Emotion Recognition System based on Right Truncated Gaussian Mixture Model" IJCA Journal, Volume 42 – issue Number 10,March- 2012

[12] V Sailaja et al "Text Independent Speaker Identification Using Finite Doubly Truncated Gaussian Mixture Model" International Journal of Information Technology and Knowledge Management, July-December 2010, Volume 2, No. 2, pp. 475-480

[13] Suri K Babu et al "Text-Independent Speaker Recognition using Emotional Features and Generalized Gamma Distribution" Volume 46 - Number 2, 2012

[14] Nagesh Vadaparthi et al "Unsupervised medical image segmentation on brain MRI images using Skew Gaussian distribution" International Conference on Recent Trends in Information Technology - ICRTIT , 2011

[15] Jianhua Tao, at al" Utterance Independent Bimodal Emotion Recognition in Spontaneous Communication" EURASIP Journal on Advances in Signal Processing,Volume-4, May 2011

An Area-Efficient Carry Select Adder Design by using 180 nm Technology

Garish Kumar Wadhwa[1]
Research Scholar, ECE
SBSSTC, Ferozepur

Amit Grover[2]
Assistant Professor, ECE
.SBSSTC, Ferozepur

Neeti Grover[3]
Assistant Professor, DASH(Poly Wing)
SBSSTC, Ferozepur

GurpreetSingh[4]
Assistant Professo
ECE, LPU, Phagwara

***Abstract*— In this paper, we proposed an area-efficient carry select adder by sharing the common Boolean logic term. After logic simplification and sharing partial circuit, we only need one XOR gate and one inverter gate in each summation operation as well as one AND gate and one inverter gate in each carry-out operation. Through the multiplexer, we can select the correct output result according to the logic state of carry-in signal. In this way, the transistor count in a 32-bit carry select adder can be greatly reduced from 1947 to 960.**

Keywords- Carry Select Adder; Area-Efficient; Hardware-Sharing; Boolean Logic

I. INTRODUCTION

The carry-ripple adder is composed of many cascaded single-bit full-adders. The circuit architecture is simple and area-efficient. However, the computation speed is slow because each full-adder can only start operation till the previous carry-out signal is ready. In the carry select adder, N bits adder is divided into M parts. Each part of adder is composed two carry ripple adders with cin_0 and cin_1, respectively. Through the multiplexer, we can select the correct output result according to the logic state of carry-in signal. The carry-select adder can compute faster because the current adder stage does not need to wait the previous stage's carry-out signal. The summation result is ready before the carry-in signal arrives; therefore, we can get the correct computation result by only waiting for one multiplexer delay in each single bit adder.

In the carry select adder, the carry propagation delay can be reduced by M times as compared with the carry ripple adder. However, the duplicated adder in the carry select adder results in larger area and power consumption. In this paper, we proposed an area-efficient carry select adder by sharing the common Boolean logic term. After Boolean simplification, we can remove the duplicated adder cells in the conventional carry select adder. Alternatively, we generate duplicate carry-out and sum signal in each single bit adder cell. By utilizing the multiplexer to select the correct output according to its previous carry-out signal, we can still preserve the original characteristics of the parallel architecture in the conventional carry select adder. In this way, the circuit area and transistor count can be greatly reduced and power delay product of the adder circuit can be also greatly lowered. The research efforts of the past years in the field of digital electronics have been directed towards the low power of digital systems. Recently, the requirement of probability and the moderate improvement in battery performance indicate power dissipation is one of the most critical design parameters day by day the demand of probability and mobility is increasing. Also the area of chip design is taken into consideration while talking about probability. Hence three most widely accepted parameters to measure the quality of a circuit or to compare various circuit styles are area, delay and power dissipation. There are three major sources of power consumption in digital CMOS circuits, which are summarized in the following equation **[I].**

$$P_{total} = P_{switching} + P_{short\text{-}circuit} + P_{leakage}$$

$$= (\alpha_{0\rightarrow1} \text{ x } C_L \text{ x } V_{dd}^2 \text{ x } f_{clk}) + (I_{sc} \text{ x } V_{dd}) + (I_{leakage} \text{ x } V_{dd})$$

The first term represents the switching component of power, where C is the load capacitance, f_{clk} is the clock frequency and $\alpha_{0\rightarrow1}$ is the node transition activity factor. The second term is due to the direct path short circuit currents, I_{sc}, which arises when both the NMOS and PMOS transistors are simultaneously active, conducting current directly from supply to ground. Finally, leakage current, $I_{leakage}$, which can arise from substrate injection and sub threshold effects, is primarily determined by fabrication technology considerations. However, while supply voltage reduction is the most effective way to reduce the power consumption, such a reduction require new design methods for low-voltage and low power integrated circuits. Since an average of 15-20% of the total power is dissipated in glitching, low power can also be achieved by reducing the glitches of the circuit [l].

II. AREA-EFFICIENT CARRY SELECT ADDER

The carry ripple adder is constructed by cascading each single-bit full-adder [1]. In the carry ripple adder, each full-adder starts its computation till previous carry-out signal is ready. Therefore, the critical path delay in a carry ripple adder is determined by its carry-out propagation path. The critical path is N-bit carry propagation path in the full-adders. As the bit number N increases, the delay time of carry ripple adder will increase accordingly in a linear way.

In order to improve the shortcoming of carry ripple adder to remove the linear dependency between computation delay time and input word length, carry select adder is presented [2]. The carry select adder divides the carry ripple adder into M parts, while each part consists of a duplicated (N/M)-bit carry ripple adder pair. This duplicated carry ripple adder pair is to anticipate both possible carry input values, where one carry ripple adder is calculated as carry input value is logic “0” and another carry ripple adder is calculated as carry input value is logic “1”. When the actual carry input is ready, either the result of carry “0” path or the result of carry “1” path is selected by the multiplexer according to its carry input value. To anticipate both possible carry input values in advance, the start of each M part carry ripple adder pair no longer need to wait for the coming of previous carry input. As a result, each M part carry ripple adder pair in the carry select adder can compute in parallel. In this way, the critical path of N bit adder can be greatly reduced.

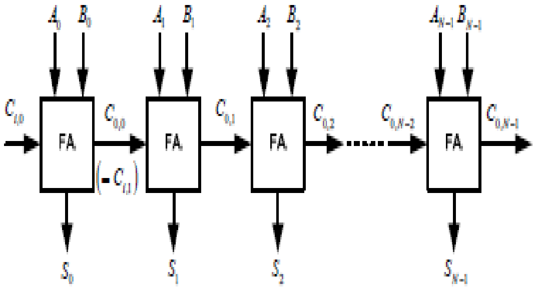


Figure 1. The N-Bit Carry Ripple Adder Constructed By N Set Single Bit Full-Adder

In the conventional N-bit carry ripple adder design, the critical path is N-bit carry propagation path plus one summation generation stage. Alternatively, the critical path is (N/M)-bit carry propagation path plus M stage multiplexer with one summation generation stage in the N-bit carry select adder. Since M is much smaller than N and delay in the multiplexer is smaller than that in the full adder, the computation delay in the carry select adder is much shorter than that in the carry ripple adder. However, implementing the adder with duplicated carry generation circuit costs almost twice hardware and twice power consumption as compared with the carry ripple adder.

Therefore, in this paper, we proposed an area-efficient carry select adder by sharing the common Boolean logic term to remove the duplicated adder cells in the conventional carry select adder. In this way, we can save many transistor counts and achieve a lower PDP As compared with the conventional carry select adder, our speed is a little slower since the parallel path in our design is shorter. However, we can achieve lower area, lower power consumption, and lower PDP. As compared with the carry ripple adder, our speed can be faster because some of the parallel architecture in the conventional carry select adder is retained. The delay time in our proposed adder design is also proportional to the bit number N; however, the delay time of multiplexer is shorter than that of full adder.

Consequently, our area-efficient adder can perform with nearly the same transistor count, nearly the same power consumption, but with faster speed and lower PDP as compared with the carry ripple adder

III. SIMULATION COMPARISON RESULTS

We compare the circuit performance with three different architectures, 32-bit carry ripple adder, 32-bit carry select adder, and 32-bit area-efficient carry select adder that is proposed in this paper. As for the transistor count, the transistor count of our proposed area-efficient carry select adder could be reduced to be very close to that of carry ripple adder; however, the transistor count in the conventional carry select adder is nearly double as compared with the proposed design. This result shows that sharing common Boolean logic term could indeed achieve a superior performance in aspect of transistor count.

The area-efficient carry select adder can also achieve an outstanding performance in power consumption. Power consumption can be greatly saved in our proposed area-efficient carry select adder because we only need one XOR gate and one INV gate in each summation operation as well as one AND gate and one OR gate in each carry-out operation after logic simplification and sharing partial circuit. Because of hardware sharing, we can also significantly reduce the occurring chance of glitch. Besides, the improvement of power consumption can be more obvious as the input bit number increases. We simulated the power consumption in the proposed area-efficient adder and the conventional carry select adder with 4, 8, 16, and 32-bit respectively in tsmc 0.18um CMOS technology. The power consumption difference between these two designs is small in the case of 4-bit input word length. Since the conventional carry select adder consists of the duplicated adder cells to prepare both the possible output values for the corresponding carry input values in advance. It not only needs larger hardware area, but also generates more glitch signals because of propagation path difference.

Therefore, as the input bit number increases, the slope of power consumption increase in the conventional carry select adder would be larger than that in our proposed design. As the input bit number of the conventional carry select adder increases to 32-bit, the power consumption in the conventional carry select adder will be 3.3 times larger than that in our proposed area-efficient carry select adder.

The conventional carry select adder performs better in terms of speed. The delay of our proposed design increases slightly because of logic circuit sharing sacrifices the length of parallel path. However, the proposed area-efficient carry select adder retains partial parallel computation architecture as the conventional carry select adder design; the delay increment of the proposed design is similar to that in the conventional design as the input bit number increases. We also simulated the delay performance in the proposed area-efficient adder and conventional carry select adder with 4, 8, 16, and 32-bit respectively.

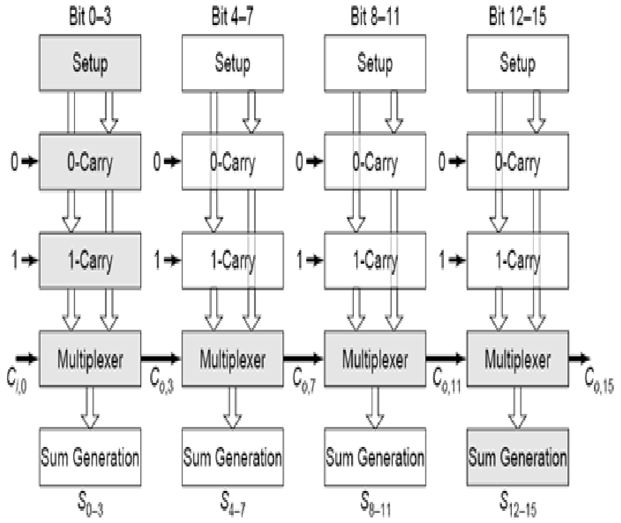


Figure 2. The 16-Bit Carry Select Adder Is Divided By The Carry Ripple Adder Into 4 Parts, While Each Part Consists Of A Duplicated 4-Bit Carry Ripple Adder Pair.

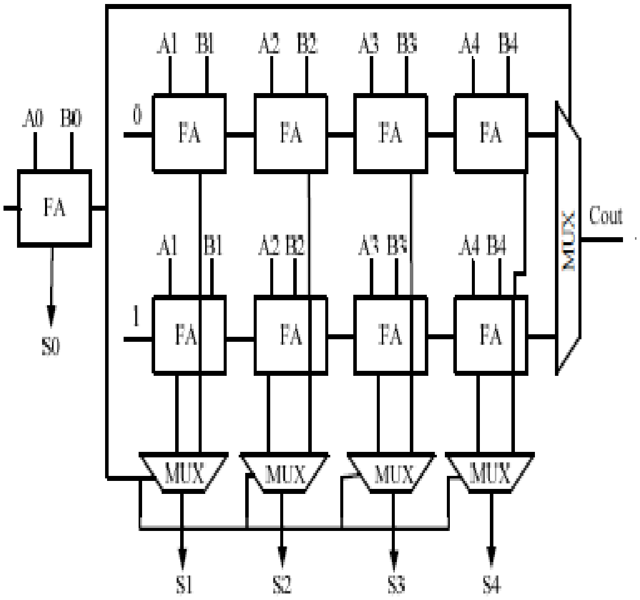


Figure 3. 5-Bit Carry Select Adder

The delay difference existing between these two designs is mainly come from the length difference in their parallel paths. In the conventional carry select adder, it divided N bits into M blocks; however, our proposed design divided every single bit as individual block. In other words, we still retain N blocks in the N bits adder. Such arrangement will lead to some speed sacrifice. We further analyze the Power-Delay-Product as shown.

The proposed area-efficient carry select adder is constructed by sharing the common Boolean logic term in summation generation**.**

We can find out that the PDP of our proposed design is smaller as compare with the conventional carry select adder and carry ripple adder design. The difference of PDP between these three designs is small in the case of the smaller input bit number.

However, as the input bits increases, the slope of power consumption increment in the conventional carry select adder would be larger than that of the proposed design. Our proposed design can compute the addition function more efficiently by means of logic circuit sharing and partial parallel computation architecture retaining; therefore, the power saving ratio in our design would be much higher than the ratio of speed sacrifice. Simplifying the carry select adder through logic simplification and partial logic circuit sharing can make the carry select adder more area-efficient and more power-efficient. The performance index of transistor count, power, delay, and PDP are summarized in Table 1.

As compared with the carry ripple adder, operation speed in our proposed carry select adder can be much faster; however, transistor count and power consumption only increase slightly. In the case of a 32-bit adder, the transistor count in the carry ripple adder is 896. The transistor count in our proposed area-efficient carry select adder is 960, which only increases 7%.

However, the transistor count in the conventional carry select adder is 1974, which increases more than twice. In terms of power consumption, we can save much power through removal of redundant logic and redundant signal switching by means of sharing common Boolean logic term. As compared with the conventional carry select adder, we can save 70% power. Relative to the carry ripple adder, we only increase 2% power. As a result, our proposed area-efficient carry select adder can perform the lowest PDP, which is only 60% of conventional carry select adder and 66% of carry ripple adder, respectively

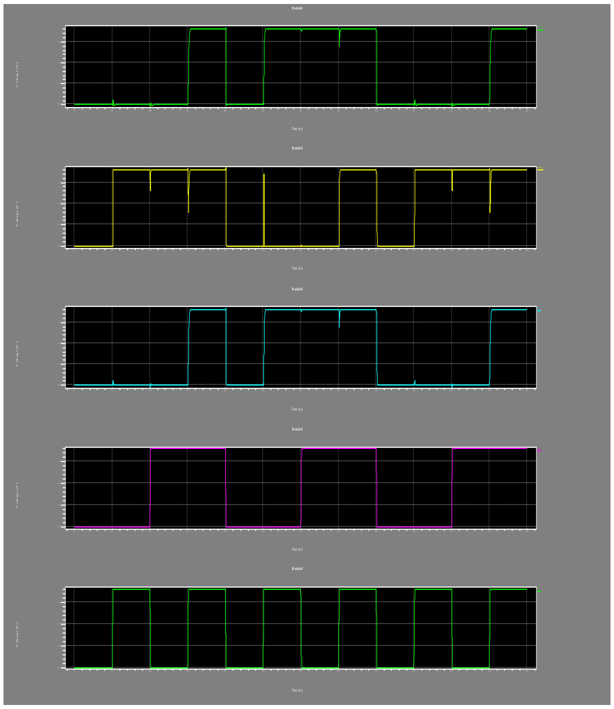
IV. RESULTS

The result is carried out at 1.8v supply voltages and average powers consumed and delay at sum and carry output are finding out

TABLE I. PERFORMANCE Parameters Of 5 Bit Carry Select Adder

Design Style	No. of transistors	Minimum Length (µm)	Avg. Power Consumption (watts)	Prop. Delay at sum (sec)	Prop. Delay at carry (sec)
Carry Select Adder	**165**	**0.18**	**3.23 x10^{-4}**	**2.15 x10^{-10}**	**4.52 x10^{-10}**

Below is the waveform of carry select adder at sum output and at carry output



V. Conclusions

In this paper, an area-efficient carry select adder is proposed. By sharing the common Boolean logic term, we can remove the duplicated adder cells in the conventional carry select adder. In this way, the transistor count in a 32-bit carry select adder can be greatly reduced from 1947 to 960. Moreover, the power consumption can be reduced from 1.26mw to 0.37mw as well as power delay product reduced from 2.14mw*ns to 1.28mw*ns. By retaining part of parallel architecture of conventional carry select adder, we can still maintain some competitiveness in speed. In this way, our area-efficient adder can perform with nearly the same transistor count, nearly the same power consumption, but with faster speed and lower PDP as compared with the carry ripple adder.

VI. Future Scope

The work can be extended to 64-bit adders. The research steps may be taken further to optimize the parameters like using frequency, capacitance, length, width etc. The work can be extended to change the technology file. The efforts can be made to decrease the transistor count so further Power, area and delay by changing the parameters. Research steps can be taken by using the other types of adders like Carry select adder, Hybrid adder etc.

References

[1] M. Rabaey, Digital Integrated Circuits," IEEE Trans. on VLSI Systems, 2003.

[2] B. Ramkumar, H.M. Kittur, and P. M. Kannan, "ASIC implementation of modified faster carry save adder," *Eur. J. Sci. Res.*, vol. 42, no. 1, pp. 53–58, 2010.

[3] T. Y. Ceiang and M. J. Hsiao, "Carry-select adder using single ripple carry adder," *Electron. Lett.*, vol. 34, no. 22, pp. 2101–2103, Oct. 1998.

[4] Y. Kim and L.-S. Kim, "64-bit carry-select adder with reduced area," *Electron. Lett.*, vol. 37, no. 10, pp. 614–615, May 2001.

[5] J. M. Rabaey, *Digtal Integrated Circuits—A Design Perspective*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[6] Y. He, C. H. Chang, and J. Gu, "An area efficient 64-bit square root carry-select adder for lowpower applications," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2005, vol. 4, pp. 4082–4085.

Author's Profile

Garish Kumar Wadhwa received his B. Tech degree in Electronics and Communication from Baba Banda Singh Bahadur Engineering College, Fatehgarh Sahib, affiliated to Punjab Technical University, Kapurthala, Punjab, India in 2010. The author place of birth is Bathinda, Punjab, India on 15th May 1988. At Present, He is doing his Master's degree in Electronics &Communication Engineering under the supervision of Mr. Amit Grover, Assistant Professor, ECE, SBSSTC, Moga road, Ferozepur-152004, Punjab, India.

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.

Neeti Grover received her master degree in Applied Sciences from Guru Nanak Dev University, Amritsar, and Punjab, India in 2007 and received her Bachelor‟s degree from Guru Nanak Dev University, Amritsar, Punjab, India in 2004. Her place of birth is Jallandhar, Punjab, India on 29th, December 1983. Currently, she is working as an Assistant Professor in the department of Applied Sciences and Humanities in Shaheed Bhagat Singh State Technical Campus (Poly Wing), Ferozpur, Punjab, India.

Gurpreet Singh The author place of birth is Faridkot, Punjab, India on 28th, August 1988. The author received M. Tech degree in Electronics and Communication Engineering from Jaypee University of Information and Technology, Solan, Himachal Pradesh, India in 2012 and received B. Tech degree in Electronics and Communication Engineering from Lovely Institutes of Technology, Phagwara, Punjab, India in 2010 with distinction. His area of interest is signal processing, MIMO Systems, Wireless mobile communications, High speed digital communicationsand4Gwireless mobile communications.

Towards No-Reference of Peak Signal to Noise Ratio

Estimation Based on Chromatic Induction Model

Jaime Moreno, Beatriz Jaime and Salvador Saucedo
Superior School of Mechanical and Electrical Engineering,
National Polytechnic Institute,
Mexico City, Mexico.

***Abstract*—The aim of this work is to define a no-referenced perceptual image quality estimator applying the perceptual concepts of the Chromatic Induction Model The approach consists in comparing the received image, presumably degraded, against the perceptual versions (different distances) of this image degraded by means of a Model of Chromatic Induction, which uses some of the human visual system properties. Also we compare our model with an original estimator in image quality assessment, PSNR. Results are highly correlated with the ones obtained by PSNR for image (99.32% *Lenna* and 96.95% for image *Baboon*), but this proposal does not need an original image or a reference one in order to give an estimation of the quality of the degraded image.**

Keywords-Human Visual System; Contrast Sensitivity Function; Perceived Images; Wavelet Transform; Peak Signal-to-Noise Ratio;No-Reference Image Quality Assessment; JPEG2000.

I. INTRODUCTION

The early years of the 21st century have witnessed a tremendous growth in the use of digital images as a means for representing and communicating information. A significant literature describing sophisticated theories, algorithms, and applications of digital image processing and communication has evolved. A considerable percentage of this literature is devoted to methods for improving the appearance of images, or for maintaining the appearance of images that are processed. Nevertheless, the quality of digital images processed or otherwise, is rarely perfect. Images are subject to distortions during acquisition, compression, transmission, processing, and reproduction. To maintain, control, and enhance the quality of images, it is important for image acquisition, management, communication, and processing systems to be able to identify and quantify image quality degradations. The development of effective automatic image quality assessment systems is a necessary goal for this purpose. Yet, until recently, the field of image quality assessment has remained in a nascent state, awaiting new models of human vision and of natural image structure and statistics before meaningful progress could be made.

Nowadays, Mean Squared Error (MSE) is still the most used quantitative performance metrics and several image quality measures are based on it, being Peak Signal-to-Noise Ratio (PSNR) the best example. But some authors like Wang and Bovik in [1], [2] consider that MSE is a poor algorithm, to be used in quality assessment systems.

Therefore it is important to know what the MSE is and what is wrong with it, in order to propose new metrics that fulfills the properties of human visual system and keeps the favorable features that the MSE has.

In this way, let $f(i,j)$ and $\hat{f}(i,j)$ represent two images being compared and the size of them is the number of intensity samples or pixels. Being $f(i,j)$ the original reference image, which has to be considered with perfect quality, and $\hat{f}(i,j)$ a distorted version of $f(i,j)$, whose quality is being evaluated. Then, the MSE and the PSNR are, respectively, defined as:

$$MSE = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}\left[f(i,j) - \hat{f}(i,j)\right]^2 \qquad (1)$$

and

$$PSNR = 10\log_{10}\left(\frac{\mathcal{G}_{max}{}^2}{MSE}\right) \qquad (2)$$

where $\mathcal{G}_{max}$ is the maximum possible intensity value in $f(i,j)$ (*M x N* size). Thus, for gray-scale images that allocate 8 bits per pixel (bpp) $\mathcal{G}_{max} = 2^8 - 1 = 255$. For color images the PSNR is defined as in the Equation 2, whereas the color MSE is the mean among the individual MSE of each component. An important task in image compression systems is to maximize the correlation among pixels, because the higher correlation at the preprocessing, the more efficient algorithm postprocessing. Thus, an efficient measure of image quality should take in to account the latter feature. In contrast to this, MSE does not need any positional information of the image, thus pixel arrangement is ordered as a one-dimensional vector. Both MSE and PSNR are extensively employed in the image processing field, since these metrics have favorable properties, such as:

- A convenient metrics for the purpose of algorithm optimization. For example in JPEG2000, MSE is used both in Optimal Rate Allocation [3], [4] and Region of interest [5], [4]. Therefore MSE can find solutions for these kind of problems, when is combined with the instruments of linear algebra, since it is differentiable.
- By definition MSE is the difference signal between the two images being compared, giving a clear meaning of the overall error signal energy.

II. IMAGE QUALITY ASSESSMENT

A. Full Reference (FR)

Bottom-Up Approaches: Psychological and physiological studies in the past century have gained us a tremendous amount of knowledge about the human visual system (HVS). Still, although much is known about the mechanisms of early, front-end vision, much more remains to be learned of the later visual pathways and the general higher level functions of the visual cortex. While the knowledge is far from complete, current models of visual information processing mechanisms have become sufficiently sophisticated that it is of interest to explore whether it is possible to deploy them to predict the performance of simple human visual behaviors, such as image quality evaluation. Bottom up approaches to image quality assessment are those methods that attempt to simulate well modeled functionalities of the HVS, and integrate these in the design of quality assessment algorithms that, hopefully, perform similar to the HVS in the assessment of image quality. In this chapter we begin with a brief description of relevant aspects of the anatomy and psychophysical features of the HVS. This description will focus on those HVS features that contribute to current engineering implementations of perceptual image quality measures. Most systems that attempt to incorporate knowledge about the HVS into the design of image quality measures use an error sensitivity framework, so that the errors between the distorted image and reference image are perceptually quantized according to HVS characteristics.

Top-Down Approaches: The bottom-up approaches to image quality assessment described in the last subsection (II-A1) attempt to simulate the functional components in the human visual system that may be relevant to image quality assessment. The underlying goal is to build systems that work in the same way as the HVS, at least for image quality assessment tasks. By contrast, the top-down systems simulate the HVS in a different way. These systems treat the HVS as a black box, and only the input output relationship is of concern. A top-down image quality assessment system may operate in a manner quite different from that of the HVS, which is of little concern, provided that it successfully predicts the image quality assessment behavior of an average human observer. One obvious approach to building such a top-down system is to formulate it as a supervised machine learning problem, as illustrated in Fig. 1. Here the HVS is treated as a black box whose inputoutput relationship is to be learned. The training data can be obtained by subjective experimentation, where a large number of test images are viewed and rated by human subjects. The goal is to train the system model so that the model prediction is minimized. This is generally a regression or function approximation problem. Many techniques are available to attack these kinds of problems.

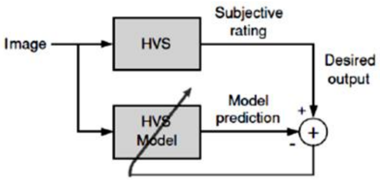


Fig. 1. Learning HVS.

Unfortunately, direct application of this method is problematic, since the dimension of the space of all images is the same as the number of pixels in the image. Furthermore, subjective testing is expensive and a typical extensive subjective experiment would be able to include only several hundred test image shardly an adequate coverage of the image space. Assigning only a single sample at each quadrant of a ten dimensional space requires a total of 1024 samples, and the dimension of the image space is in the order of thousands to millions; An excellent example of the problem of dimensionality.

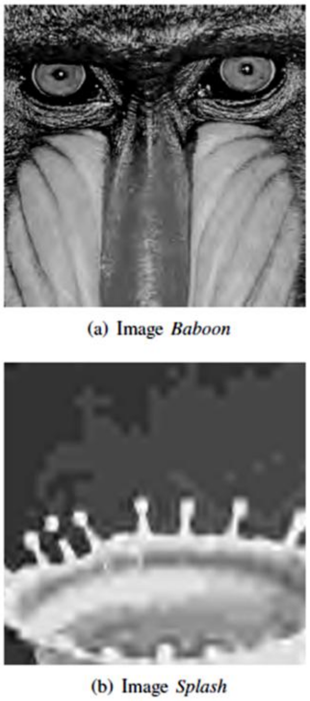
One method that might be useful to overcome this problem is by dimension reduction. The idea is to map the entire image space onto a space of much lower dimensionality by exploiting knowledge of the statistical distribution of typical images in the image space. Since natural images have been found to exhibit strong statistical regularities, it is possible that the cluster of typical natural images may be represented by a low dimensional manifold, thus reducing the number of sample images that might be needed in the subjective experiments.

However, dimension reduction is no trivial task. Indeed, no dimension reduction technique has been developed to reduce the dimension of natural images to 10 or less (otherwise, extremely efficient image compression techniques would have been proposed on the basis of such reduction). Consequently, using a dimension reduction approach for general purpose image quality assessment remains quite difficult. Nonetheless, such an approach may prove quite effective in the design of application specific quality assessment systems, where the types of distortions are fixed and known and may be described by a small number of parameters.

B. No-Reference (NR)

No-reference (NR) image quality assessment is, perhaps, the most difficult (yet conceptually simple) problem in the field of image analysis. By some means, an objective model must evaluate the quality of any given real world image, without referring to an original high quality image. On the surface, this seems to be a mission impossible. How can the quality of an image be quantitatively judged without having a numerical model of what a good/bad quality image is supposed to look like? Yet, amazingly, this is quite an easy task for human observers. Humans can easily identify high quality images versus low quality images, and, furthermore, they are able to point out what is right and wrong about them without seeing the original. Moreover, humans tend to agree with each other to a pretty high extent. For example, without looking at the original image, probably every reader would agree that the noisy, blurry, and JPEG2000 compressed images in Fig. 2 have lower quality than the luminance shifted and contrast stretched images.

Before developing any algorithm for image quality assessment, a fundamental question that must be answered is what source of information can be used to evaluate the quality of images. Clearly, the human eyebrain system is making use of a very substantial and effective pool of information about images in making subjective judgments of image quality.



(a) Image *Baboon*

(b) Image *Splash*

Fig. 2. 256 x 256 patches (cropped for visibility) of Images *Baboon* and *Splash* distorted by means of JPEG2000 compression, although both images have the same objective quality (PSNR=30dB), their visual quality is very different.

Three types of knowledge may be employed in the design of image quality measures: knowledge about the original high quality image, knowledge about the distortion process, and knowledge about the human visual system (HVS). In FR quality assessment, the high quality original image is known a priori. In NR quality assessment, however, the original image is absent, yet one can still assume that there exists a high quality original image, of which the image being evaluated is a distorted representation. It is also reasonable to make a further assumption that such a conjectured original image belongs to the set of typical natural images.

It is important to realize that the cluster of natural images occupies an extremely tiny portion in the space of all possible images. This potentially provides a strong prior knowledge about what these images should look like. Such prior knowledge could be a precious source of information for the design of image quality measures. Models of such natural scenes attempt to describe the class of high quality original images statistically. Interestingly, it has been long conjectured in computational neuroscience that the HVS is highly adapted to the natural visual environment, and that, therefore, the modeling of natural scenes and the HVS are dual problems.

Knowledge about the possible distortion processes is another important information source that can be used for the development of NR image quality measures. For example, it is known that blur and noise are often introduced in image acquisition and display systems and reasonably accurate models are sometimes available to account for these distortions. Images compressed using block based algorithms such as JPEG often exhibit highly visible and undesirable blocking artifacts. Wavelet based image compression algorithms operating at low bit rates can blur images and produce ringing artifacts near discontinuities. Of course, all of these types of distortions are application dependent. An application specific NR image quality assessment system is one that is specifically designed to handle a specific artifact type, and that is unlikely to be able to handle other types of distortions. The question arises, of course, whether an application specific NR system is truly reference free, since much information about the distorted image is assumed. However, nothing needs to be assumed about the original image, other than, perhaps models derived from natural scene statistics or other natural assumptions. Since the original images are otherwise unknown, we shall continue to refer to more directed problems such as these as application specific NR image quality assessment problems.

Of course, a more complex system that includes several modes of artifact handling might be constructed and that could be regarded as approaching general purpose NR image quality assessment. Before this can happen, however, the various components need to be designed. Fortunately, in many practical application environments, the distortion processes involved are known and fixed. The design of such application specific NR quality assessment systems appears to be much more approachable than the general, assumption free NR image quality assessment problem. Very little, if any, meaningful progress has been made on this latter problem. Owing to a paucity of progress in other application specific areas, this work mainly focuses on NR image quality assessment methods, which are designed for assessing the quality of compressed images. In particular, attention is given to a spatial domain method and a frequency domain method for block based image compression, and a wavelet domain method for wavelet based image compression.

III. The *NR*PSNR Algorithm

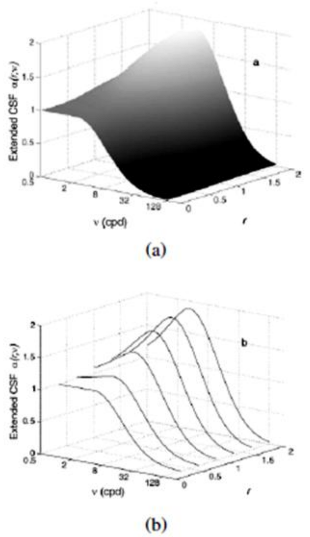
A. *Chromatic Induction Wavelet Model (CIWaM)*

The Chromatic Induction Wavelet Model (CIWaM) [6] is a low-level perceptual model of the HVS. It estimates the image perceived by an observer at a distance d just by modeling the perceptual chromatic induction processes of the HVS. That is, given an image $\mathcal{I}$ and an observation distance d, CIWaM obtains an estimation of the perceptual image $\mathcal{I}_\rho$ that the observer perceives when observing $\mathcal{I}$ at distance d. CIWaM is based on just three important stimulus properties: spatial frequency, spatial orientation and surround contrast. These three properties allow unifying the chromatic assimilation and contrast phenomena, as well as some other perceptual processes such as saliency perceptual processes [7].

The CIWaM model takes an input image $\mathcal{I}$ and decomposes it into a set of wavelet planes $\omega_{s,o}$ of different spatial scales *s* (i.e., spatial frequency ν) and spatial orientations *o*. It is described as:

$$\mathcal{I} = \sum_{s=1}^{n} \sum_{o=v,h,dgl} \omega_{s,o} + c_n , \qquad (3)$$

where *n* is the number of wavelet planes, c_n is the residual plane and *o* is the spatial orientation either *v*ertical, *h*orizontal or *d*ia*g*ona*l*.



(a)

(b)

Fig. 3. (a) Graphical representation of the e-CSF for the luminance channel. (b) Some profiles of the same surface along the Spatial Frequency axis for different center-surround contrast energy ratio values (r). The psychophysically measured CSF is a particular case of this family of curves (concretely for r = 1).

The perceptual image $\mathcal{I}_\rho$ is recovered by weighting these $\omega_{s,o}$ wavelet coefficients using the *extended Contrast Sensitivity Function* (e-CSF, Fig. 3). The e-CSF is an extension of the psychophysical CSF [8] considering spatial surround information (denoted by *r*), visual frequency (denoted by ν, which is related to spatial frequency by observation distance) and observation distance (d). Perceptual image $\mathcal{I}_\rho$ can be obtained by

$$\mathcal{I}_\rho = \sum_{s=1}^{n} \sum_{o=v,h,dgl} \alpha(\nu, r)\, \omega_{s,o} + c_n \; , \qquad (4)$$

where $\alpha(\nu, r)$is the e-CSF weighting function that tries to reproduce some perceptual properties of the HVS. The term $\alpha(\nu, r) \;\; \omega_{s,o} \;\; \equiv \;\; \omega_{s,o;\rho,d}$ can be considered the *perceptual wavelet coefficients* of image $\mathcal{I}$ when observed at distance *d* and is written as:

$$\alpha(\nu, r) = z_{ctr} \cdot C_d(\dot{s}) + C_{min}(\dot{s}) \; . \qquad (5)$$

This function has a shape similar to the e-CSF and the three terms that describe it are defined as:

z_{ctr} Non-linear function and estimation of the central feature contrast relative to its surround contrast, oscillating from zero to one, defined by:

$$z_{ctr} = \frac{\left[\frac{\sigma_{cen}}{\sigma_{sur}}\right]^2}{1 + \left[\frac{\sigma_{cen}}{\sigma_{sur}}\right]^2} \qquad (6)$$

being σ_{cen} and σ_{sur} the standard deviation of the wavelet coefficients in two concentric rings, which represent a center-surround interaction around each coefficient.

$C_d(\dot{s})$ Weighting function that approximates to the perceptual e-CSF, emulates some perceptual properties and is defined as a piecewise Gaussian function [8], such as:

$$C_d(\dot{s}) = \begin{cases} e^{-\frac{\dot{s}^2}{2\sigma_1^2}}, & \dot{s} = s - s_{thr} \leq 0, \\ e^{-\frac{\dot{s}^2}{2\sigma_2^2}}, & \dot{s} = s - s_{thr} > 0. \end{cases} \qquad (7)$$

$C_{min}(\dot{s})$ Term that avoids $\alpha(\nu, r)$ function to be zero and is defined by:

$$C_{min}(\dot{s}) = \begin{cases} \frac{1}{2}\, e^{-\frac{\dot{s}^2}{2\sigma_1^2}}, & \dot{s} = s - s_{thr} \leq 0, \\ \frac{1}{2}, & \dot{s} = s - s_{thr} > 0. \end{cases} \qquad (8)$$

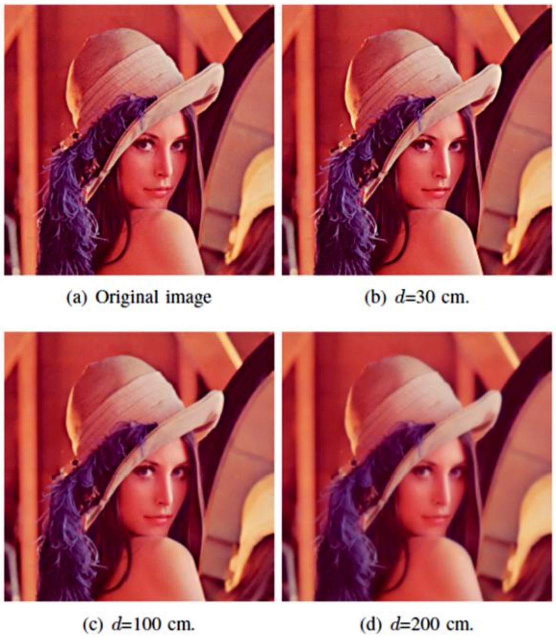
taking $\sigma_1 = 2$ and $\sigma_2 = 2\sigma_1$. Both $C_{min}(\dot{s})$ and $C_d(\dot{s})$ depend on the factor s_{thr}, which is the scale associated to 4 cycles per degree when an image is observed from the distance *d* with a pixel size $l_p$and one visual degree, whose expression is defined by Equation 9. Where s_{thr} value is associated to the e-CSF maximum value

$$s_{thr} = \log_2 \left(\frac{d \tan(1^\circ)}{4\, l_p} \right) \qquad (9)$$

Fig. 4 shows three examples of CIWaM images of *Lenna*, calculated by Eq. 4 for a 19 inch monitor with 1280 pixels of horizontal resolution, at *d* = {30, 100, 200} centimeters.

B. Basics

In the no-referenced image quality issue, there is only a distorted version $\hat{f}(i,j) \;=\; \Lambda[f(i,j)]$ that is compared with $f(i,j)$, being Λ a distortion model and the unknown original image $f(i,j)$ is considered a pattern Υ ([0,1;1,0]) like a chessboard (Figs. 5) with the same size of $\hat{f}(i,j)$.



(a) Original image

(b) *d*=30 cm.

(c) *d*=100 cm.

(d) *d*=200 cm.

Fig. 4. (a) Original color image *Lenna* . (b)-(d) Perceptual images obtained by CIWaM at different observation distances *d*.

The difference between these two images depends on the features of the distortion model Λ. For example, blurring, contrast change, noise, JPEG blocking or JPEG2000 wavelet ringing.

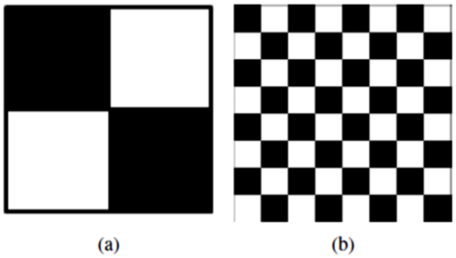


Fig. 5. (a) Pattern [0,1;1,0] or Υ . (b) Pattern Υ repeated sixteen times.

In Fig. 2, the images *Babbon* and *Splash* are compressed by means of JPEG2000. These two images have the same PSNR=30 dB when compared to their corresponding original image, that is, they have the same numerical degree of distortion (i.e. the same objective image quality PSNR). But, their subjective quality is clearly different, showing the image *Baboon* a better visual quality. Thus, for this example, PSNR and perceptual image quality has a small correlation. On the image *Baboon*, high spatial frequencies are dominant. A modification of these high spatial frequencies by Λ induces a high distortion, resulting a lower PSNR, even if the modification of these high frequencies are not perceived by the HVS. In contrast, on image *Splash*, mid and low frequencies are dominant. Modification of mid and low spatial frequenciesalso introduces a high distortion, but they are less perceived by the HVS. Therefore, correlation of PSNR against the opinion of an observer is small. Fig. 6 shows the diagonal high spatial frequencies of these two images, where there are more high frequencies in image *Baboon*.

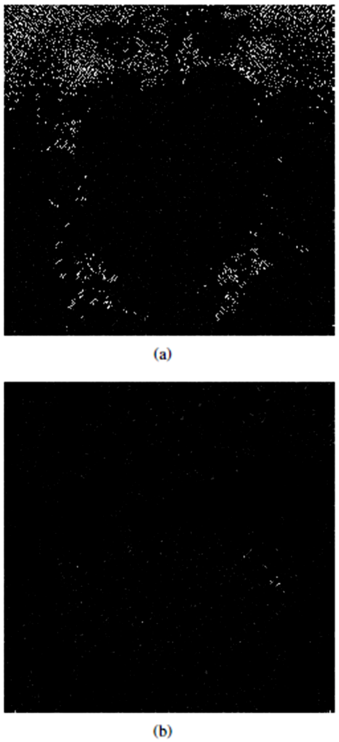


Fig. 6. Diagonal spatial orientation of the first wavelet plane of Images (a) *Baboon* and (b)*Splash* distorted by JPEG2000 with PSNR=30dB.

If a set of distortions $\hat{f}_k(i,j) = \Lambda_k[f(i,j)]$ is generated and indexed by *k* (for example, let _ be a blurring operator), the image quality of $\hat{f}_k(i,j)$evolves while varying *k*, being *k*, for example, the degree of blurring. Hence, the evolution of $\hat{f}_k(i,j)$ depends on the characteristics of the original $f(i,j)$. Thus, when increasing *k*, if $f(i,j)$ contains many high spatial frequencies the PSNR rapidly decreases, but when low and mid frequencies predominated PSNR slowly decreases.

Similarly, the HVS is a system that induces a distortion on the observed image $\hat{f}_k(i,j)$, whose model is predicted by CIWaM. Hence, CIWaM is considered a HSV particular distortion model $\Lambda \equiv \mathrm{CIWaM}$ that generates a perceptual image $\hat{f}_\rho(i,j) \equiv \mathcal{I}_\rho$ from an observed image $f(i,j) \equiv \mathcal{I}$, i.e $\mathcal{I}_\rho = CIWaM[\mathcal{I}]$. Therefore, a set of distortions is defined as $\Lambda_k \equiv \mathrm{CIWaM}_d$, being *d* the observation distance. That is, a set of perceptual images is defined $\mathcal{I}_{\rho,d} = \mathrm{CIWaM}_d[\mathcal{I}]$which is considered a set of perceptual distortions of the hypothetical image $\mathcal{I}$.

When image $\hat{f}_k(i,j)$ is observed at distance $\bar{d}$ and this distance is reduced, the artifacts, if this possesses, are better perceived. In contrast, $\hat{f}_k(i,j)$is observed from a far distance human eyes cannot perceive their artifacts, in consequence, the perceptual image quality of the distorted image is always high. The distance where the observer can perceive the best image quality of image $\hat{f}_k(i,j)$is considered as the distance *D*.

Let $f(i,j)$ and $\hat{f}(i,j)$ be an pattern image and a distorted image, respectively. NRPSNR methodology is based on finding a distance D, where there is no perpetual difference between the wavelet energies of the images $f(i,j)$and $\hat{f}(i,j)$, when an observer observe them at d centimeters of observation distance. So measuring the PSNR of $\hat{f}(i,j)$at D will yield a fairer and No-reference perceptual evaluation of its image quality.

*NR*PSNR algorithm is divided in five steps, which is summarized by the Figure 7 and described as follows:

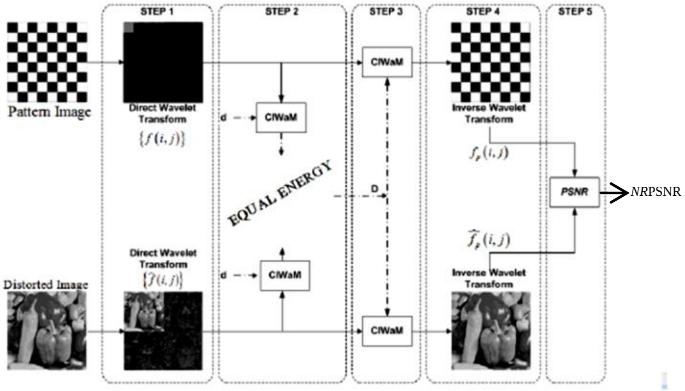


Fig. 7. Methodology for No-Reference PSNR weighting by means of CIWaM. Both Pattern and Distorted images are wavelet transformed. The distance *D* where the energy of perceptual images obtained by CIWaM are equal is found. Then, PSNR of perceptual images at *D* is calculated, obtaining the *NR*PSNR metrics.

Step 1: Wavelet Transformation

Forward wavelet transform of images $f(i,j)$ and $\hat{f}(i,j)$ is performed using Eq. 3, obtaining the sets $\{\omega_{s,o}\}$ and $\{\hat{\omega}_{s,o}\}$, respectively. The employed analysis filter is the Daubechies 9-tap/7-tap filter (Table I).

TABLE I. 9/7 ANALYSIS FILTER.

Analysis Filter		
i	Low-Pass Filter $h_L(i)$	High-Pass Filter $h_H(i)$
0	0.6029490182363579	1.115087052456994
±1	0.2668641184428723	-0.5912717631142470
±2	-0.07822326652898785	-0.05754352622849957
±3	-0.01686411844287495	0.09127176311424948
±4	0.02674875741080976	

Step 2: Distance D

The total energy measure or the *deviation signature*[9] $\bar{\varepsilon}$ is the absolute sum of the wavelet coefficient magnitudes, defined by [10]

$$\bar{\varepsilon} = \sum_{n=1}^{N} \sum_{m=1}^{M} |x(m,n)| \tag{10}$$

where *x*(*m; n*) is the set of wavelet coefficients, whose energy is being calculated, being *m* and *n* the indexes of the coefficients. Basing on the traditional definition of a calorie, the units of $\bar{\varepsilon}$ are wavelet calories (wCal) and can also be defined by Eq. 10, since one wCal is the energy needed to increase the absolute magnitude of a wavelet coefficient by one scale.

From wavelet coefficients $\{\omega_{s,o}\}$ and $\{\hat{\omega}_{s,o}\}$ the corresponding perceptual wavelet coefficients $\left\{\omega_{s,o;\rho,\tilde{d}}\right\} = \alpha(\nu,r)$. $\omega_{s,o}$ and $\left\{\hat{\omega}_{s,o;\rho,\tilde{d}}\right\} = \alpha(\nu,r) \cdot \hat{\omega}_{s,o}$ are obtained by applying CIWaM with an observation distance $\tilde{d}$. Therefore, Equation 11 expresses the relative wavelet energy ratio $\varepsilon\mathcal{R}\left(\tilde{d}\right)$, which compares how different are the energies of the reference and distorted CIWaM perceptual images, namely ε_ρ and $\widehat{\varepsilon}_\rho$ respectively, when these images are watched from a given distance $\tilde{d}$.

$$\varepsilon\mathcal{R}\left(\tilde{d}\right) = 10 \cdot \left|\log_{10} \frac{\varepsilon_\rho\left(\tilde{d}\right)}{\widehat{\varepsilon}_\rho\left(\tilde{d}\right)}\right| \tag{11}$$

Thus, the main goal of this step is to find $\varepsilon\mathcal{R}(D)$, namely, at *D* ε_ρ is equal to $\widehat{\varepsilon}_\rho$, where the energy of the distorted images are the same than the energy of the pattern.

Step 3: Perceptual Images

Getting the perceptual images $\{f_p(i,j)\}$and $\{\hat{f}_p(i,j)\}$ from the $\{f(i,j)\}$ and $\{\hat{f}(i,j)\}$images watched at *D* centimeters, using Equation 4.

Step 4: Inverse Wavelet Transformation

Perform the Inverse Wavelet Transform of $\{\omega_{s,o;\rho,D}\}$ and $\{\hat{\omega}_{s,o;\rho,D}\}$, obtaining the perceptual images $f_{\rho(i,j),D}$ and $\hat{f}_{\rho(i,j),D}$, respectively. The synthesis filter in Table II is an inverse Daubechies 9-tap/7-tap filter.

TABLE II. 9/7 SYNTHESIS FILTER.

Synthesis Filter		
i	Low-Pass Filter $h_L(i)$	High-Pass Filter $h_H(i)$
0	1.115087052456994	0.6029490182363579
±1	0.5912717631142470	-0.2668641184428723
±2	-0.05754352622849957	-0.07822326652898785
±3	-0.09127176311424948	0.01686411844287495
±4		0.02674875741080976

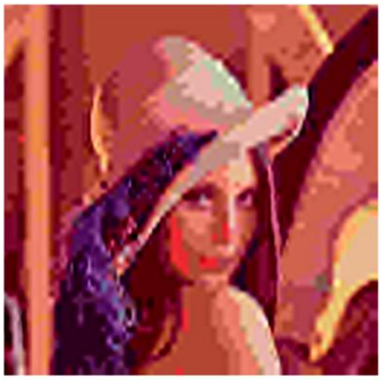
Step 5:* PSNR *between perceptual images

Calculate the PSNR between perceptual images $f_{\rho(i,j),D}$ and $\hat{f}_{\rho(i,j),D}$ using Eq. 2 in order to obtain the No-Reference CIWaM weighted PSNR i.e. the *NR*PSNR.

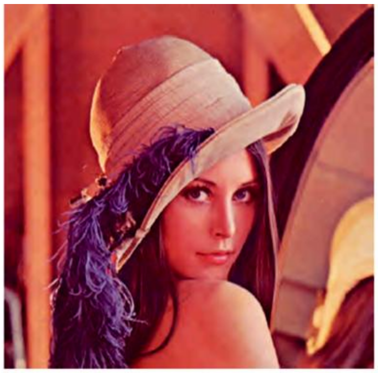
IV. Experimental Results

It is important to mention that *NR*PSNR estimates the degradation, thus, the smaller the better. In this section, we indistinctly use either *NR*PSNR or BPSNR, since *NR*PSNR is the blind version of PSNR, thus, *NR*PSNR performance is assessed by comparing the statistical significance of the images *Lenna* and *Baboon*, in addition to the Pearson correlation between *NR*PSNR and PSNR data.

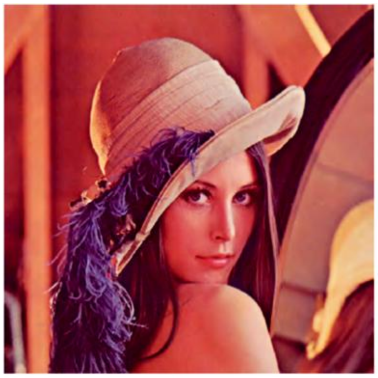
Figure 8 depicts three JPEG2000 distorted versions of the image *Lenna* with 0.05(Fig. 8(a)), 0.50 (Fig. 8(b)) and 1.00 (Fig. 8(c)) bits per pixel. PSNR estimates 23.41, 32.74 and 34.96 dB, respectively. While *NR*PSNR computes 48.42, 36.56 and 35.95 dB, respectively. Thus, both PSNR and *NR*PSNR estimate that image at 1.00 bpp has lower distortion.



(a) 0.05 bpp



(b) 0.50 bpp



(c) 1.00 bpp

Fig. 8. JPEG2000 Distorted versions of color image *Lenna* at different bit rates expressed in bits per pixel (bpp). (a) High Distortion, (b) medium Distortion and (c) Low Distortion.

Figure 10 depicts three JPEG2000 distorted versions of the image *Baboon* with 0.05(Fig. 10(a)), 0.50 (Fig. 10(b)) and 1.00 (Fig. 10(c)) bits per pixel. PSNR estimates 18.55, 23.05 and 25.11 dB, respectively. While *NR*PSNR computes 43.49, 30.07 and 28.71 dB, respectively.

Thus, both PSNR and *NR*PSNR estimate that image at 0.05 bpp has higher distortion. When this experiment is extended computing the JPEG2000 distorted versions from 0.05 bpp to 3.00bpp (increments of 0.05 bpp, depicted at Figure 11), we found that the correlation between PSNR and *NR*PSNR is 96.95 %, namely for image *Baboon* for every 10,000 estimation *NR*PSNR misses only in 305 assessments.

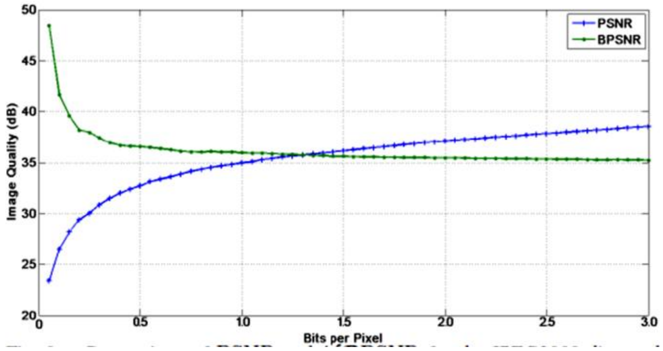
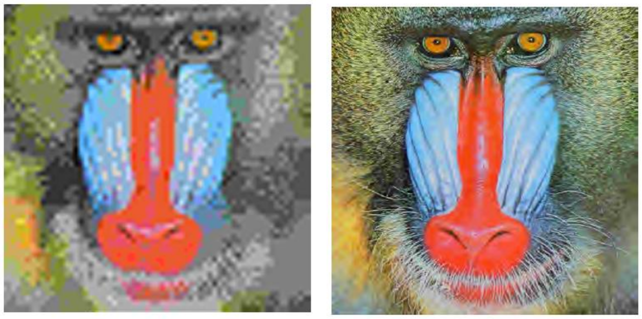
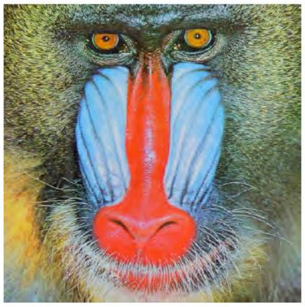


Fig. 9. Comparison of PSNR and *NR*PSNR (Blind-PSNR or BPSNR) for the JPEG2000 distorted versions of image *Lenna*.



(a) 0.05 bpp

(b) 0.50 bpp



(c) 1.00 bpp

Fig. 10. JPEG2000 Distorted versions of color image *Baboon* at different bit rates expressed in bits per pixel (bpp). (a) High Distortion, (b) Medium Distortion and (c) Low Distortion.

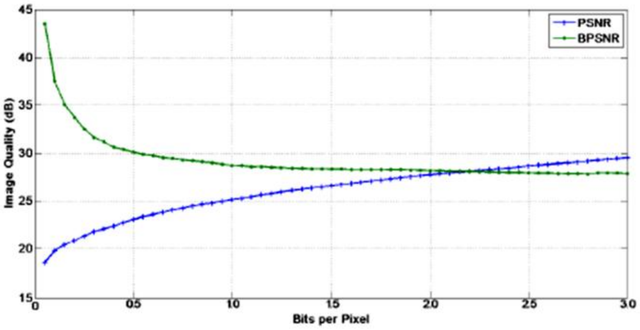


Fig. 11. Comparison of PSNR and *NR*PSNR (Blind-PSNR or BPSNR) for the JPEG2000 distorted versions of image *Baboon*

V. CONCLUSIONS

*NR*PSNR is a new metric for no-reference or blind image quality based on perceptual weighting of PSNR by using a perceptual low-level model of the Human Visual System (CIWaM model). The proposed *NR*PSNR metrics is based on five steps.

The *NR*PSNR assessment was tested in two well-known images, such as *Lenna* and *Baboon*. It is a well-correlated image quality method in these images for JPEG2000 distortions when compared to PSNR. Concretely, *NR*PSNR correlates with PSNR, on the average in 98.13%. It is possible to quantize a particular pixel while an algorithm of bit allocation is working, incorporating into embedded compression schemes such as EZW, SPIHT, JPEG2000 or H*i*-SET[11].

ACKNOWLEDGMENT

This work is supported by The National Polytechnic Institute of Mexico by means of a granted fund by the Committee of Operation and Promotion of Academic Activities (COFAA).

REFERENCES

[1] Z. Wang and A. Bovik, “Mean squared error: Love it or leave it? a new look at signal fidelity measures,” *Signal Processing Magazine, IEEE*, vol. 26, no. 1, pp. 98 –117, jan. 2009.

[2] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment*, 1st ed. Morgan & Claypool Publishers: Synthesis Lectures on Image, Video, & Multimedia Processing, February 2006.

[3] F. Auli-Llinas and J. Serra-Sagrista, “Low complexity JPEG2000 rate control through reverse subband scanning order and coding passes concatenation,” *IEEE Signal Processing Letters*, vol. 14, no. 4, pp. 251 –254, april 2007.

[4] S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, ser. ISBN: 0-7923-7519-X. Kluwer Academic Publishers, 2002.

[5] J. Bartrina-Rapesta, F. Auli-Llinas, J. Serra-Sagrista, and J. Monteagudo-Pereira, “JPEG2000 Arbitrary ROI coding through rate-distortion optimizationtechniques,” in *Data Compression Conference*, 25-27 2008, pp. 292 –301.

[6] X. Otazu, C. P´arraga, and M. Vanrell, “Toward a unified chromatic induction model,” *Journal of Vision*, vol. 10(12), no. 6, 2010.

[7] N. Murray, M. Vanrell, X. Otazu, and A. Parraga, “Saliency estimation using a non-parametric low-level vision model,” in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR’2011)*, 2010, pp. 433 –440.

[8] K. T. Mullen, “The contrast sensitivity of human colour vision to redgreen and blue-yellow chromatic gratings.” *The Journal of Physiology*, vol. 359, pp. 381–400, February 1985.

[9] G. van de Wouwer, P. Scheunders, and D. van Dyck, “Statistical texture characterization from discrete wavelet representations,” *IEEE Transactions on Image Processing*, vol. 8, no. 4, pp. 592 –598, Apr. 1999.

[10] B. A. Wilson and M. A. Bayoumi, “A computational kernel for fast and efficient compressed-domain calculations of wavelet subband energies.” *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 50, no. 7, pp. 389 – 392, July 2003.

[11] J. Moreno and X. Otazu, “Image coder based on H*i*lbert Scaning of Embedded quadTrees,” *IEEE Data Compression Conference*, p. 470, March 2011.

A Block Cipher Involving a Key and a Key Bunch Matrix, Supplemented with 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 block cipher involving a key and a key bunch matrix. In this cipher, we have made use of key-based permutation and key-based substitution. The cryptanalysis carried out in this investigation, shows very clearly, that this cipher is a very strong one. This is all on account of the confusion and the diffusion created by the permutation, the substitution, in each round of the iteration process.**

Keywords-Key; key bunch matrix; encryption; decryption; permutation; substitution; avalanche effect; cryptanalysis

I. Introduction

The study of the block ciphers [1] is an interesting area of research in cryptography. In the very recent past, we have developed a pair of block ciphers [2-3], which include a key matrix, as it is in the case of the Hill cipher, and a key bunch matrix. In these investigations, we have made use of the concepts of the modular arithmetic inverse and the multiplicative inverse.

In [2], we have made use of function Mix(), which mixes the binary bits in each round of the iteration process, and in [3], we have introduced a function called Permute(), which carries out permutation of binary bits of the plaintext in each round of the iteration process. In these analyses, we have noticed that the key matrix and the key bunch matrix, and the additional function Mix()/ Permute() strengthen the cipher, in a conspicuous manner.

In the present paper, our objective is to develop a block cipher, wherein we use a key matrix together with a key bunch matrix. Here, we have introduced a key-based permutation and a substitution basing upon the key. In this, our interest is to see, how the permutation and the substitution would influence the cipher and enhance the strength of the cipher, due to the confusion and the diffusion arising in this process.

We now mention the plan of the paper. In section 2, we discuss the development of the cipher and introduce the flowcharts and the algorithms required in this analysis. We illustrate the cipher and discuss the avalanche effect in section 3. We study the cryptanalysis in section 4. Finally in section 5, we deal with the computation carried out in this investigation and draw conclusions.

II. Development of the Cipher

We consider a plain P having n(2) characters and represent it in the form of a square matrix of size n by using EBCDIC code. Thus we have

$$P = [\ p_{ij}\ ],\ i=1 \text{ to } n,\ j=1 \text{ to } n. \qquad (2.1)$$

Let the key matrix K be given by

$$K=[\ k_{ij}\ ],\ i=1 \text{ to } n,\ j=1 \text{ to } n, \qquad (2.2)$$

The encryption key bunch matrix E is taken in the form

$$E = [\ e_{ij}\ ],\ i=1 \text{ to } n,\ j=1 \text{ to } n, \qquad (2.3)$$

wherein each e_{ij} is an odd number lying in [1-255].

On using the concept of the multiplicative inverse [4], the decryption key bunch matrix D is obtained in the form

$$D= [dij],\ i=1 \text{ to } n,\ j=1 \text{ to } n, \qquad (2.4)$$

It is to be noted her that all the elements of D are also odd numbers which lie in [1-255].

The basic equations governing the encryption can be written in the form

$$P = (KP) \bmod 256, \qquad (2.5)$$
$$P = [\ e_{ij} \times p_{ij}\ ] \bmod 256,\ i=1 \text{ to } n,\ j = 1 \text{ to } n \qquad (2.6)$$
$$P = \text{Permute}(P), \qquad (2.7)$$
$$P= \text{Substitute}(P), \qquad (2.8)$$
and
$$C = P. \qquad (2.9)$$

The corresponding equations of the decryption process are given by

$$C = \text{ISubstitute}(C) \qquad (2.10)$$
$$C = \text{IPermute}(C), \qquad (2.11)$$
$$C = [\ d_{ij} \times c_{ij}\ ] \bmod 256,\ i=1 \text{ to } n,\ j = 1 \text{ to } n, \qquad (2.12)$$
$$C = (K(-1)\ C) \bmod 256, \text{ and} \qquad (2.13)$$
$$P = C. \qquad (2.14)$$

The details of the function Permute() and the function Substitute() are explained later. It is to be noted here, that the functions ISubstitute() and IPermute() denote the reverse process of the functions Substitute() and Permute().

The flowcharts depicting the process of the encryption and the decryption are given in Figs. 1 and 2.

The algorithms, for the encryption and the decryption are as follows.

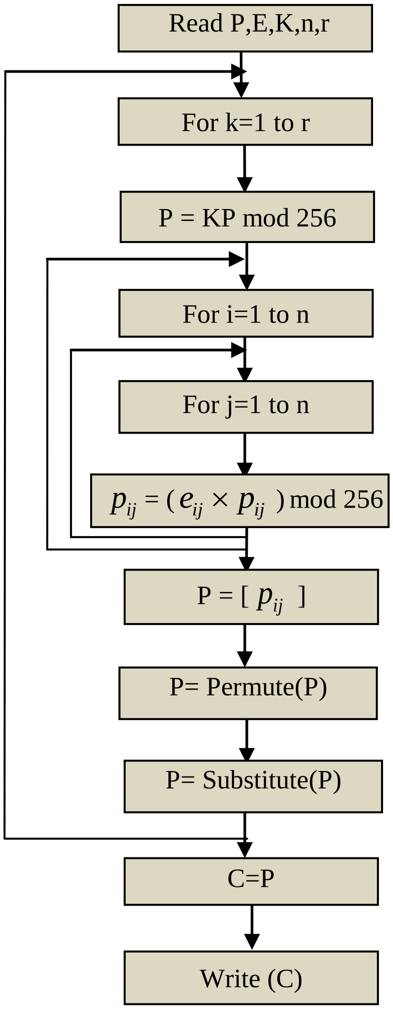


Fig.1 Flowchart for Encryption

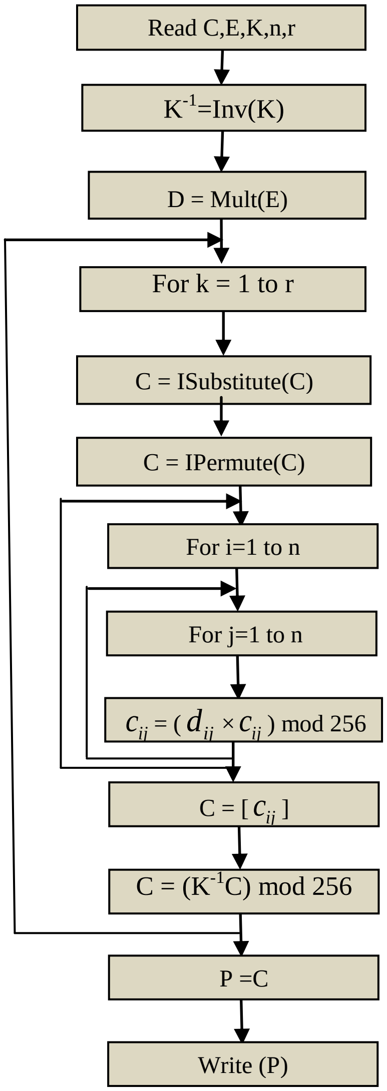


Fig.2. Flowchart for Decryption

Algorithm for Encryption

```
1.  Read P,E,K,n,r
2.  For k = 1 to r do
    {
3.  P=(KP) mod 256
4.  For i=1 to n do
    {
5.  For j=1 to n do
    {
6.  $p_{ij} = (e_{ij} \times p_{ij})$ mod 256
    }
    }
7.  P=[ $p_{ij}$ ]
8.  P=Permute(P)
9.  P=Substitute(P)
    }
8.  C=P
9.  Write(C)
```

Algorithm for Decryption

```
1.  Read C,E,K,n,r
2.  $K^{-1}$=Inv(K)
3.  D=Mult(E)
4.  For k = 1 to r do
    {
5.  C=ISubstitute(C)
6.  C=IPermute(C)
7.  For i =1 to n do
    {
8.  For j=1 to n do
    {
9.  $c_{ij} = (d_{ij} \times c_{ij})$ mod 256
```

```
        }
        }

        }
  12. P=C
  13. Write (P)
```

Let us now, explain the basic ideas underlying in functions Permute() and Substitute(). Both are dependent on a key. Let us take the key K in the form

$$K = \begin{bmatrix} 120 & 182 & 102 & 13 \\ 25 & 14 & 16 & 200 \\ 30 & 147 & 61 & 122 \\ 40 & 127 & 206 & 91 \end{bmatrix} \qquad (2.15)$$

The numbers in this key are listed in the 2nd row of the following table, Table-1.

```
  10. C=[ $c_{ij}$ ]
  11. C = ($K^{-1}$C) mod 256
```

$$\begin{bmatrix} p_{111}p_{112}\cdot\cdot p_{118} & p_{121}p_{122}\cdot\cdot p_{128} \\ p_{211}p_{212}\cdot\cdot p_{218} & p_{221}p_{222}\cdot\cdot p_{228} \\ p_{311}p_{312}\cdot\cdot p_{318} & p_{321}p_{322}\cdot\cdot p_{328} \\ p_{411}p_{412}\cdot\cdot p_{418} & p_{421}p_{422}\cdot\cdot p_{428} \\ p_{131}p_{132}\cdot\cdot p_{138} & p_{141}p_{142}\cdot p_{148} \\ p_{231}p_{232}\cdot\cdot p_{238} & p_{241}p_{242}\cdot p_{248} \\ p_{331}p_{332}\cdot\cdot p_{338} & p_{341}p_{342}\cdot p_{348} \\ p_{431}p_{432}\cdot\cdot p_{438} & p_{441}p_{442}\cdot p_{448} \end{bmatrix} \qquad (2.19)$$

TABLE-1. RELATION BETWEEN SERIAL NUMBERS AND THE ASCENDING ORDER OF THE KEY NUMBERS.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
120	182	102	13	25	14	16	200	30	147	61	122	40	127	206	91
10	14	9	1	4	2	3	15	5	13	7	11	6	12	16	8

The 1st row of this table contains the serial number, and the 3rd row of this table indicates the ascending order of the numbers in the key, given in the 2nd row.

Consider the plaintext P in any round of the iteration process. It is possible to see this plaintext as a set of square matrices of size 16 whenever n is divisible by 16. As the Table-1 is suggesting, we interchange the rows

(1,10), (2,14), (3,9), (5,4), (8,15), (11,7) and (13,6).
(2.16)

Similarly, it may be done in the case of the columns. It may be noted here, that once we have made an interchange involving a row or a column, we do not do anymore interchange involving that row or column subsequently so that plaintext remains in a systematic manner. This is the basic idea underlying in the function Permute(), when n>=16. On the other hand, when n takes n takes a value less than 16, for example when n=4, then let us see how the process of the permutation will be carried out.

Consider an example when n=4. In this case, the plaintext is of the form

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} \qquad (2.17)$$

On representing each element of this matrix in terms of binary bits, in a row-wise manner, we have This is a matrix having 4 rows and 32 columns. This can be written, for convenience, in the form of another matrix, given by (2.19).

This matrix has 8 rows and 16 columns. In order to carry out permutation, we swap the rows (5,4) as indicated by (2.16). The rest of the rows are untouched, as we do not have the possibility of interchange. Then the columns are interchanged by following the content of (2.16).

Let us now consider the process of substitution, which depends upon the permutation. The EBCDIC code, which includes the number 0 to 255, can be written in the form of a matrix, given by

$$EB(i, j) = [16(i-1) + j - 1], i = 1\, to\, n,\ j = 1\, to\, n, \qquad (2.20)$$

This has 16 rows and 16 columns. On interchanging the rows first and then the columns next, we get a new matrix, having the numbers 0 to 255, in some other order. This table can be written in the form, given in (2.21).

On noting the correspondence between the matrices, given by (2.20) and (2.21), we can perform the substitution process in any plaintext. Thus we have the function Substitute().

The function Inv() is used to obtain the modular arithmetic inverse of the key matrix K. The function Mult() results in the decryption key bunch matrix D for the given encryption key bunch matrix E. For a thorough understanding of these

$$\begin{bmatrix} p_{111}p_{112}\cdot\cdot p_{118} & p_{121}p_{122}\cdot\cdot p_{128} & p_{131}p_{132}\cdot\cdot p_{138} & p_{141}p_{142}\cdot p_{148} \\ p_{211}p_{212}\cdot\cdot p_{218} & p_{221}p_{222}\cdot\cdot p_{228} & p_{231}p_{232}\cdot\cdot p_{238} & p_{241}p_{242}\cdot p_{248} \\ p_{311}p_{312}\cdot\cdot p_{318} & p_{321}p_{322}\cdot\cdot p_{328} & p_{331}p_{332}\cdot\cdot p_{338} & p_{341}p_{342}\cdot p_{348} \\ p_{411}p_{412}\cdot\cdot p_{418} & p_{421}p_{422}\cdot\cdot p_{428} & p_{431}p_{432}\cdot\cdot p_{438} & p_{441}p_{442}\cdot p_{448} \end{bmatrix} \qquad (2.18)$$

$$SB=\begin{bmatrix} 153 & 157 & 152 & 148 & 147 & 156 & 154 & 158 & 146 & 144 & 150 & 155 & 149 & 145 & 151 & 159 \\ 217 & 221 & 216 & 212 & 211 & 220 & 218 & 222 & 210 & 208 & 214 & 219 & 213 & 209 & 215 & 223 \\ 137 & 141 & 136 & 132 & 131 & 140 & 138 & 142 & 130 & 128 & 134 & 139 & 133 & 129 & 135 & 143 \\ 73 & 77 & 72 & 68 & 67 & 76 & 74 & 78 & 66 & 64 & 70 & 75 & 69 & 65 & 71 & 79 \\ 57 & 61 & 56 & 52 & 51 & 60 & 58 & 62 & 50 & 48 & 54 & 59 & 53 & 49 & 55 & 63 \\ 201 & 205 & 200 & 196 & 195 & 204 & 202 & 206 & 194 & 192 & 198 & 203 & 197 & 193 & 199 & 207 \\ 169 & 173 & 168 & 164 & 163 & 172 & 170 & 174 & 162 & 160 & 166 & 171 & 165 & 161 & 167 & 175 \\ 233 & 237 & 232 & 228 & 227 & 236 & 234 & 238 & 226 & 224 & 230 & 235 & 229 & 225 & 231 & 239 \\ 41 & 45 & 40 & 36 & 35 & 44 & 42 & 46 & 34 & 32 & 38 & 43 & 37 & 33 & 39 & 47 \\ 9 & 13 & 8 & 4 & 3 & 12 & 10 & 14 & 2 & 0 & 6 & 11 & 5 & 1 & 7 & 15 \\ 105 & 109 & 104 & 100 & 99 & 108 & 106 & 110 & 98 & 96 & 102 & 107 & 101 & 97 & 103 & 111 \\ 185 & 189 & 184 & 180 & 179 & 188 & 186 & 190 & 178 & 176 & 182 & 187 & 181 & 177 & 183 & 191 \\ 89 & 93 & 88 & 84 & 83 & 92 & 90 & 94 & 82 & 80 & 86 & 91 & 85 & 81 & 87 & 95 \\ 25 & 29 & 24 & 20 & 19 & 28 & 26 & 30 & 18 & 16 & 22 & 27 & 21 & 17 & 23 & 31 \\ 121 & 125 & 120 & 116 & 115 & 124 & 122 & 126 & 114 & 112 & 118 & 123 & 117 & 113 & 119 & 127 \\ 249 & 253 & 248 & 244 & 243 & 252 & 250 & 254 & 242 & 240 & 246 & 251 & 245 & 241 & 247 & 255 \end{bmatrix} \quad (2.21)$$

functions, we may refer to [2].

In this cipher, r denotes the number of rounds carried out in the iteration process. Here, we have taken r=16.

III. Illustration of the Cipher and the Avalanche Effect

Consider the plaintext given below.

Dear Brother! With all the training that you had from NCC in your college, having a strong feel that India is our motherland, and it is our responsibility to protect this country from the invasion by other countries, you left us some years back. After that, there are several changes within the country. When you were leaving us, we had only few parties such as Congress, Communist and BJP. Today, parties have grown as mushrooms and the number of parties is that many. We do not know, in what way unity can be achieved in this country! Each party wants to destroy the other party, each party want to come to power, and each thinks that it must rule the whole country, crushing all the other parties. Ethical values have gone down! Each person want to earn crores and crores, so that he would be able to build up his own party, and to feed all the members entering into his party in a grand manner with additional facilities, such as liquor and all the other attractions satisfying the passion. This is the fate of the country! You may protect6 the country at the borders, but I do not know who can protect this country within this country from the tyranny of all the political parties and the people supporting them. (3.1)

On focusing our attention on the first 16 characters, we have

Dear Brother! Wi (3.2)

On using the EBCDIC code, we write the plaintext (3.2) in the form

$$P=\begin{bmatrix} 196 & 133 & 129 & 153 \\ 64 & 194 & 153 & 150 \\ 163 & 136 & 133 & 153 \\ 79 & 64 & 230 & 137 \end{bmatrix} \quad (3.3)$$

Let us take the key matrix K, in the form

$$K=\begin{bmatrix} 120 & 182 & 102 & 13 \\ 25 & 14 & 16 & 200 \\ 30 & 147 & 61 & 122 \\ 40 & 127 & 206 & 91 \end{bmatrix} \quad (3.4)$$

Here, it may be noted that we have taken this K as the same as (2.15), as this is having modular arithmetic inverse.

Let us take E in the form

$$E=\begin{bmatrix} 121 & 157 & 11 & 239 \\ 11 & 167 & 189 & 23 \\ 167 & 105 & 17 & 19 \\ 237 & 109 & 33 & 187 \end{bmatrix} \quad (3.5)$$

On using the plaintext P, the key matrix K, the encryption key bunch matrix E, given by (3.3) – (3.5), and applying the encryption algorithm, we get the ciphertext C in the form

$$C=\begin{bmatrix} 116 & 103 & 184 & 219 \\ 174 & 112 & 253 & 194 \\ 231 & 86 & 28 & 189 \\ 239 & 198 & 119 & 132 \end{bmatrix} \quad (3.6)$$

On using the concept of multiplicative inverse, we get the decryption key bunch matrix D in the form

$$D=\begin{bmatrix} 201 & 181 & 163 & 15 \\ 163 & 23 & 149 & 167 \\ 23 & 217 & 241 & 27 \\ 229 & 101 & 225 & 115 \end{bmatrix} \quad (3.7)$$

On using the C, the D, and the K, given by (3.6), (3.7) and (3.4), and applying the decryption algorithm, we get back the plaintext P.

Let us now examine the avalanche effect. On replacing the 4th row 4th column element, 137 by 153, we have a change of one binary bit in the plaintext P. On using this modified plaintext, the K, and the E, and employing the encryption algorithm, we get the ciphertext C in the form

$$C = \begin{bmatrix} 147 & 104 & 57 & 131 \\ 21 & 46 & 177 & 26 \\ 8 & 46 & 235 & 121 \\ 5 & 197 & 189 & 55 \end{bmatrix} \qquad (3.8)$$

On comparing (3.6) and (3.8), after putting them in their binary form, we find that these two ciphertexts differ by 70 bits out of 128 bits.

Let us now consider a one binary bit change in the key K. On replacing the 2nd row 3rd column element, 16 of the key K, given by (3.4), by 48, we have a one bit change. On using this modified key, the plaintext P, and the encryption key bunch matrix E, and the encryption algorithm, given in section 2, the ciphertext corresponding to the modified key is obtained in the form

$$C = \begin{bmatrix} 141 & 180 & 117 & 2 \\ 255 & 166 & 5 & 61 \\ 158 & 130 & 243 & 140 \\ 94 & 20 & 136 & 3 \end{bmatrix} \qquad (3.9)$$

On comparing (3.6) and (3.9), after putting them in their binary form, we notice that these two ciphertexts differ by 81 bits out of 128 bits.

From the above discussion, we conclude that, this cipher exhibits a strong avalanche effect, which stands as a benchmark in respect of the strength of the cipher.

IV. Cryptanalysis

This is the analysis which enables us to establish the strength of the cipher. The different types of attacks available in the literature of the cryptography are

1. Ciphertext only attack (Brute force attack),
2. Known plaintext attack,
3. Chosen plaintext attack, and
4. Chosen ciphertext attack.

Generally, an analytical proof is offered in the first two cases, and a checkup is done with all possible intuitive ideas in the latter two cases. A cipher is said to be acceptable, if it withstands the first two attacks [1].

In this cipher, we are having a key matrix K and key bunch matrix E. Both are taken to be square matrices of size n. In view of this fact, the size of the key space is

$$2^{8n^2} \times 2^{7n^2} = 2^{15n^2} = \left(2^{10}\right)^{1.5n^2} \approx \left(10^{3}\right)^{1.5n^2} = 10^{4.5n^2}. \qquad (4.1)$$

$$\frac{10^{4.5n^2} \times 10^{-7}}{365 \times 24 \times 60 \times 60} = 3.12 \times 10^{4.5n^2 - 15} \ \textit{years}. \qquad (4.2)$$

On assuming that the time required for the computation with one value of the key and the one value of the E, in the key space as 10^{-7}, the time required for the execution of the cipher with all possible keys (i.e., taking all possible pairs of K and E, into consideration) in the key space is

Specifically, in this analysis, as we have n=4, the time given by (4.2), takes the form 3.12 x 10^{57} years. As this time is very large, we conclude that, this cipher cannot be broken by the brute force attack.

Let us examine the known plaintext attack. Here, we have as many pairs of plaintexts and ciphertexts that we like to have, can be had, at our disposal. Confining our attention to r=1, that is to only one round of the iteration process, the system of equations governing the encryption process, can be written in the form

P = (KP) mod 256, (4.3)

P = [$e_{ij} \times p_{ij}$] mod 256, i=1 to n, j = 1 to n, (4.4)

P = Permute(P), (4.5)

P = Substitute(P), (4.6)

and

C = P. (4.7)

From (4.7), we can readily have P, as we know C. on using this P, we cannot proceed further, from bottom, as the function Substitute() and ISubstitute() depend upon the key K. Though P on the right hand of (4.3) is known to us, we cannot proceed further, as the P on the left hand side of (4.3) is unknown. In view of the above facts, we cannot the break this cipher by the known plaintext attack.

As the equations, governing the encryption process, are found to be very much involved, in view of the functions Permute() and Substitute(), which are based upon the key, and the modulo arithmetic operation, we cannot imagine to choose, intuitively, any plaintext or ciphertext, for breaking the cipher.

In the light of the above discussion, we conclude that this cipher cannot be broken by any attack, and it is a strong one by all means.

V. Computations and Conclusions

In this investigation, we have developed a block cipher which involves the basic ideas of the Hill cipher [5] and the basic concepts of the key bunch matrix. Here, we have made use of the functions Permute() and Substitute(), for permuting the plaintext and for modifying the plaintext, by the substitution process.

On account of these functions and the iteration process, the plaintext has undergone several modifications, in the process of encryption.

The programs required for carrying out the encryption and the decryption are written in Java.

The plaintext, given by (3.1), is divided into 77 blocks. As the last block is having only 2 characters, we have added 14 zeroes as additional characters to make it a complete block. On carrying out the encryption of each block separately, by using the K and the E, we get the ciphertext corresponding to the entire plaintext (3.1), in the form (5.1). The cryptanalysis carried out in this investigation, has clearly shown that this cipher is strong one and it cannot be broken by any attack. This investigation can be modified by including a large size key matrix and a corresponding encryption key bunch matrix. Then this can be applied to the encryption of images and security of images can be achieved very conveniently.

REFERENCES

[1] William Stallings: Cryptography and Network Security: Principle and Practices”, Third Edition 2003, Chapter 2, pp. 29.

[2] Dr. V.U.K. Sastry, K.Shirisha, “□A Block Cipher Involving a Key Matrix and a Key bunch Matrix, Supplemented with Mix”, in press.

[3] Dr. V.U.K. Sastry, K.Shirisha, “□A Block Cipher Involving a Key Matrix and a Key bunch Matrix, Supplemented with Permutation”, in The International Journal of Engineering And Science (IJES), ISSN: 2319 – 1813 ISBN: 2319 – 1805, Vol. – No.2, Dec 2012, pp. 40-47.

[4] Dr. V.U.K. Sastry, K.Shirisha, “A Novel Block Cipher Involving a Key Bunch Matrix”, in International Journal of Computer Applications (IJCA) (0975 – 8887) Vol.55– No.16, Oct 2012, Foundation of Computer Science, NewYork, pp. 1-6.

[5] Lester Hill, (1929), “Cryptography in an algebraic alphabet”, V.36 (6), pp. 306-312., American Mathematical Monthly.

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. She is pursuing her Ph.D. Her research interests are Information Security and Data Mining. She published 9 research papers in International Journals. She stood University topper in the M.Tech.(CSE).

116	103	184	219	174	112	253	194	231	86	28	189	239	198	119	132
37	56	27	189	138	59	142	125	58	185	212	103	214	143	253	180
64	201	234	6	44	136	105	138	207	151	84	125	1	131	62	131
136	242	189	159	246	47	225	142	183	148	126	51	60	160	111	88
146	136	229	20	75	75	6	240	64	235	223	233	208	44	181	44
133	231	147	4	191	52	202	99	89	93	15	22	103	83	32	13
17	152	58	92	210	103	15	1	161	111	82	31	24	16	217	241
53	165	124	62	45	253	33	234	219	179	9	254	240	161	149	122
101	239	9	82	187	206	184	13	113	242	65	102	9	97	88	213
250	68	21	136	162	212	90	218	140	214	31	77	7	90	109	111
154	83	27	33	116	197	130	250	207	135	232	125	11	212	31	112
40	135	34	255	187	46	110	43	193	218	20	246	200	93	31	103
246	210	248	224	120	28	23	184	201	244	125	146	75	241	182	77
132	108	32	215	59	163	141	116	231	80	38	51	135	246	115	39
211	218	200	35	146	77	191	167	207	252	34	148	37	89	223	134
179	112	71	157	174	36	177	154	9	221	103	33	82	42	166	78
173	50	94	118	237	187	166	168	52	104	8	186	0	32	16	106
254	121	1	122	41	17	98	186	123	113	206	133	112	249	205	73
42	79	88	23	20	183	56	16	220	36	211	143	230	18	75	242
105	6	237	70	68	162	186	42	192	212	23	55	15	239	44	33
99	136	5	76	223	227	156	239	70	130	240	68	170	148	249	60
11	200	206	192	111	89	254	25	155	105	216	228	181	91	31	212
8	78	2	32	133	153	58	213	222	49	204	51	103	16	110	214
114	162	135	26	101	39	49	155	73	104	4	80	149	85	128	102
178	99	204	157	29	150	220	247	90	65	77	170	132	241	9	204
237	179	19	35	173	183	33	168	96	34	125	68	183	190	55	153
197	154	204	185	180	240	25	85	135	166	37	16	80	152	218	161
219	20	16	78	62	237	82	47	40	150	228	217	93	61	73	233
139	214	34	105	102	95	209	242	110	135	243	22	221	19	3	31
105	136	83	15	114	55	190	114	44	31	203	204	143	155	110	247
15	122	13	155	234	107	10	53	159	69	17	55	108	18	49	66
43	187	158	57	214	151	111	210	135	206	53	168	172	57	212	71
247	152	148	234	53	103	133	156	149	55	108	143	118	155	41	62
39	252	244	247	28	71	5	193	186	20	197	147	115	41	191	46
53	139	75	148	68	165	113	134	12	198	123	105	66	238	59	36
122	39	156	228	48	249	183	80	99	33	143	49	197	138	83	214
203	139	3	190	174	248	69	10	145	27	228	166	158	254	98	186
29	14	244	67	252	102	28	21	203	243	186	13	217	173	79	124
46	186	241	83	116	241	208	95	133	88	97	60	19	100	44	207
22	253	173	139	97	128	26	211	240	175	33	16	189	62	224	179
255	200	28	179	212	140	178	106	183	6	206	231	245	101	139	248
206	30	203	17	84	60	238	2	172	203	157	30	87	181	58	199
181	216	58	204	177	224	9	5	43	93	23	153	105	15	122	96
73	234	221	206	228	250	183	101	134	88	209	98	26	129	24	96
6	173	185	245	95	18	116	53	145	113	133	58	78	144	88	195
217	33	97	223	90	182	21	75	175	135	148	8	178	97	208	101
153	69	41	98	121	227	146	190	220	41	114	203	126	13	235	143
70	153	196	174	111	123	205	193	66	53	217	53	189	245	218	147
73	96	114	251	47	122	221	80	53	120	232	79	160	146	140	59
77	125	111	197	156	102	86	22	29	184	210	172	243	121	221	181
12	139	124	223	119	208	241	168	162	213	228	214	103	80	227	102
136	230	96	102	107	153	58	198	107	174	82	216	190	37	125	55
104	176	1	189	115	72	73	200	201	209	250	243	255	80	172	7

$$\begin{matrix}
81 & 32 & 59 & 178 & 183 & 217 & 133 & 179 & 231 & 203 & 207 & 128 & 246 & 119 & 81 & 62 \\
55 & 61 & 56 & 229 & 42 & 33 & 145 & 112 & 82 & 243 & 229 & 126 & 185 & 149 & 154 & 38 \\
92 & 126 & 1 & 136 & 223 & 37 & 76 & 248 & 61 & 38 & 85 & 234 & 124 & 163 & 133 & 106 \\
76 & 229 & 178 & 120 & 38 & 189 & 141 & 139 & 164 & 128 & 48 & 30 & 49 & 107 & 154 & 26 \\
194 & 159 & 88 & 12 & 45 & 6 & 212 & 105 & 9 & 218 & 86 & 107 & 178 & 254 & 72 & 52 \\
16 & 252 & 20 & 174 & 80 & 61 & 3 & 121 & 72 & 185 & 57 & 193 & 97 & 80 & 174 & 113 \\
47 & 105 & 22 & 82 & 5 & 140 & 4 & 99 & 112 & 201 & 149 & 172 & 141 & 95 & 127 & 65 \\
111 & 226 & 137 & 109 & 93 & 208 & 77 & 110 & 223 & 240 & 103 & 187 & 25 & 0 & 66 & 54 \\
87 & 70 & 124 & 16 & 161 & 250 & 155 & 78 & 172 & 166 & 184 & 203 & 237 & 155 & 138 & 162 \\
48 & 173 & 149 & 227 & 17 & 171 & 17 & 252 & 241 & 71 & 114 & 211 & 234 & 26 & 109 & 233 \\
34 & 203 & 112 & 156 & 80 & 156 & 30 & 222 & 29 & 211 & 154 & 233 & 121 & 142 & 244 & 226 \\
109 & 103 & 255 & 214 & 126 & 112 & 82 & 54 & 206 & 44 & 164 & 111 & 38 & 50 & 170 & 181 \\
169 & 152 & 20 & 34 & 52 & 205 & 196 & 16 & 249 & 125 & 127 & 173 & 148 & 140 & 182 & 100 \\
29 & 137 & 247 & 206 & 198 & 170 & 147 & 143 & 97 & 145 & 182 & 180 & 19 & 152 & 76 & 140 \\
11 & 215 & 73 & 3 & 12 & 6 & 227 & 237 & 113 & 28 & 80 & 107 & 152 & 108 & 226 & 49 \\
72 & 81 & 236 & 101 & 99 & 81 & 121 & 222 & 114 & 252 & 41 & 131 & 60 & 145 & 42 & 164 \\
136 & 151 & 255 & 122 & 178 & 149 & 94 & 18 & 9 & 49 & 52 & 52 & 83 & 179 & 176 & 137 \\
88 & 201 & 184 & 97 & 101 & 202 & 46 & 47 & 40 & 200 & 12 & 208 & 197 & 205 & 110 & 51 \\
28 & 134 & 28 & 94 & 118 & 6 & 165 & 100 & 14 & 98 & 125 & 124 & 130 & 168 & 228 & 214 \\
139 & 249 & 61 & 52 & 60 & 199 & 54 & 210 & 225 & 238 & 68 & 8 & 105 & 151 & 143 & 138 \\
119 & 183 & 186 & 132 & 219 & 30 & 232 & 244 & 171 & 153 & 187 & 93 & 104 & 37 & 22 & 33 \\
241 & 210 & 23 & 176 & 205 & 20 & 107 & 66 & 126 & 17 & 208 & 219 & 16 & 73 & 232 & 224 \\
180 & 15 & 226 & 39 & 119 & 139 & 167 & 146 & 51 & 89 & 53 & 187 & 67 & 30 & 199 & 99 \\
234 & 6 & 59 & 199 & 6 & 195 & 55 & 195 & 162 & 132 & 151 & 63 & 168 & 62 & 14 & 84
\end{matrix} \tag{5.1}$$

Semantic E-Learn Services and Intelligent Systems using Web Ontology

K.Vanitha, K.Yasudha
Assistant Professor,
Department of Computer Science,
GITAM University, Visakhapatnam.

Dr.M.Sri Venkatesh
Associate Professor,
Department of Computer Science,
GITAM University, Visakhapatnam

K.N.Sowjanya
Assistant Professor
Dept.IT,
GITAM University

***Abstract*-----Present vision for the web is the semantic web in which information is given explicit meaning, making it easier for machines to automatically process and integrate information available on the web. It provides the information exactly. Now days, ontology is playing a major role in knowledge representation for the semantic web [1]. Ontology is a conceptualization of domain into a human understandable and machine readable or machine process able format consisting of entities, attributes, relationships and axioms. Ontology web language is designed for use by applications that need to process the content of information [22]. In this context many e-learning systems were proposed in the literature. Semantic Web technology may support more advanced Artificial intelligence problems for knowledge retrieval [20]. This paper aims at presenting an intelligent e-learning system from the literature.**

Keywords-Semantic web; e-learning; Ontology Web Language (OWL); Ontology; OWL-S Service Ontology.

I. INTRODUCTION

The emergence of web technologies for data and knowledge interaction gives rise to the need for supportive frameworks for knowledge distribution. Semantic web in which information is given explicit meaning, making it easier for machines to automatically process and integrate information available on the web aimed at providing shared semantic spaces for web contents[12]. Now days with the rapid development of technology the learning methods have been changed. E-learning systems are taking prominent role in making the humans learning methods apart from the class room teaching irrespective of their age, income etc., in this scenario, in the literature there are many methods have been proposed and used [3]. Fayed et al proposed a model based on semantic web technology which is used by the Qatar university students and faculty of engineering [2]. Another intelligent web teacher system for learning personalization using semantic web model was proposed by Nicola, Gaeta1 [3] and there is an adaptive educational hypermedia systems [AEHS] by Metteo et al. This paper aims at presenting intelligent e-learning systems modeled by Fayed et.al and Nicola et al and Mateo et al.

II. SEMANTIC WEB

In recent years Semantic Web is the hottest topic in the area of AI and in the internet community. Semantic Web performs the meaning (semantics) of information and services on the web, and making it possible for the web to "understand" and satisfy the requests of people and machines to use the web content which is the idea of world wide web inventor Tim Berners-Lee. Semantic web builds an appropriate infrastructure for intelligent agents to verify the web, while performing complex actions for their users. Ultimately, Semantic Web is about how to implement reliable, large-scale interoperation of Web services, to make such services computer interpretable – to create a Web of machine-understandable and interoperable services that intelligent agents can discover, execute and compose automatically [2].

The latest view of the semantic web has been changed as services. These services can be divided on two families "world services" and "web Services".

The example for a world service includes a shop, a museum, a restaurant, whose address type and description is accessible over the web. In contrast, a web service is a resource that can be automatically retrieved and invoked over the web [11]. Web service based applications can consider as conglomerates of independent, autonomous services developed by independent parties. Such components are not integrated at design time; they are integrated dynamically at runtime according to the current needs [15]. For example, an e-learning course can be assembled dynamically by composing learning objects stored in independent repositories.

A. Meta data

The preliminary source for performing semantic web operations is based on metadata. Metadata is "data about data". The aim of incorporating the Meta data is to find the data sources from the web, when end-user tries to search for information on the web [11]. Generally the data sources will be heterogeneous which belongs to different types i.e., unstructured, semi-structured and structured. Generally for the semantic web the data source will be a document, a web page, textual content, data, audio or video [8].

In the Semantic web, documents are marked up with semantic metadata which is machine-understandable about the human readable content of documents. The following are the different types for Meta data.

- Syntactic Metadata: The simplest form of metadata which describes non-contextual information about content and provides general information.

- Structural Metadata: Provides the information regarding the structure of the content and describes how items are arranged.
- Semantic Metadata: This adds relationships, rules, and constraints to syntactic and structural metadata and describes contextually relevant or domain-specific information about content based on ontology [21].

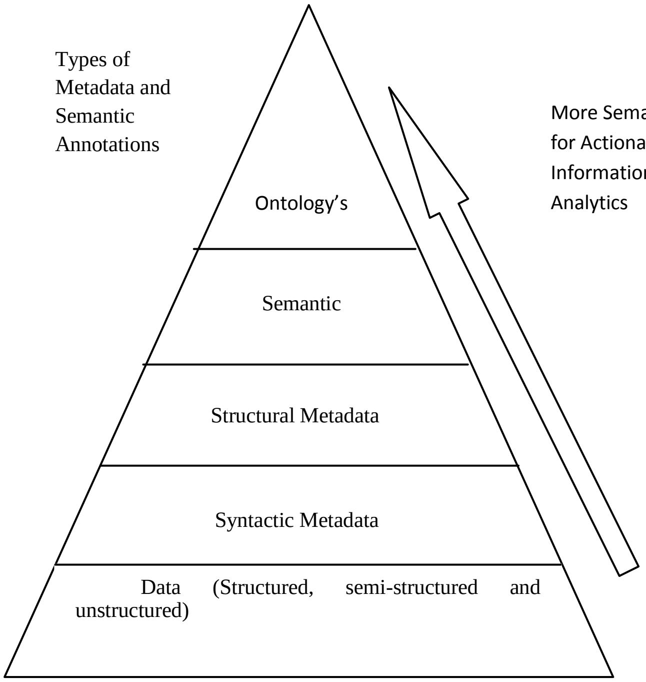

Types of Metadata and Semantic Annotations
More Semantics for Actionable Information and Analytics
Ontology's
Semantic
Structural Metadata
Syntactic Metadata
Data (Structured, semi-structured and unstructured)

Fig1: Types of Metadata

A. *OWL-S Service Ontology*

OWL-D is an OWL service upper ontology that offers a Vocabulary that can be used in conjunction with OWL to describe services in an unambiguous, computer interpretable format. OWL-S was developed with the goal of allowing discovery, invocation, composition, and automatic monitoring of Web services (Martin et al, 2006). OWL-S treats service composition as processes. There is a very clear distinction among process properties, Structure, and implementation in OWL-S, which provides a way to model a process independently of its implementation.

The web service technology will revolutionize the way software is developed. Some of the potential benefits of the web services technologies are decentralization , speed, software packing and the other extreme web service technology has received a deal of criticism for providing an over simplified model . It leads out several fundamental concepts as Data definition, service invocation behavior mediation, composition and service guarantees.

The technology will allow a distributed and decentralized way of web services [11]. A positive effect of the increase of transactions through the web is forcing to adapt a more dynamic and user centered service model.

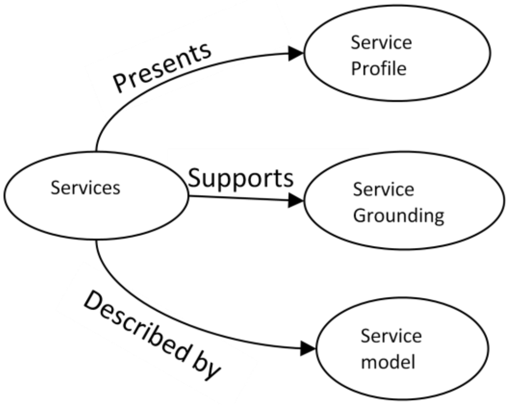


Fig:2 OWL-S Service Ontology

It is transforming response time into the competitive advantage. The web service compositional model has the potential to review the format and allow to be developed as service components. In over simplified model of concepts there are no domain specific data definitions. It is used to model the input and output of every application that is depending upon application domain.

III. Web Ontology

Ontology is about the exact description of things and their relationships. Ontology's are considered one of the pillars of the Semantic Web; although they do not have a universally accepted definition According to Tom Gruber [17] ontology is a formal specification of a shared conceptualization [18]. For the web, ontology is about the exact description of web information and relationships between web information. The purpose of the Web Ontology domain is to be able to model the relationships between prominent web ontology's and map them onto equivalent freebase types and topics.

IV. An Adaptive Educational Hypermedia System (AEHS)

The focus of Mateo et al. is on the aspects of personalization. They proposed a model as "An Adaptive Educational Hypermedia System" which supports the individual in the process of finding, selecting, accessing and retrieving web resources [2].

This model is based on the concepts of adaptive hypermedia system [19]. This adaptive hyper media system is in turn based on hypermedia system which was presented in brief in this paper.

A. *Personalization*

The goal of personalization in the Semantic web is to make easier the access to the right resources. This task entitles two processes [19] [5]. They are retrieval and presentation. Retrieval consists in finding or constructing the right resources when they are needed, either on demand otherwise, when the information arises in the work [8]. Personalization is a process of filtering the access to web content according to the individual needs and requirements of each particular user.

B. Adaptive Hypermedia System

This enumerates the functionality of a hypermedia system which personalizes for the individual users.

C. Hypermedia System

A hypermedia system consists of documents which are connected by links[6]. Thus, there are mainly two aspects which can be adapted to the users: the content and the links.

D. Content Level

There are five methods identified for content level adaption.

- Additional explanation method which displays those parts of a document fits to user goals ,interest, tasks, knowledge etc.,
- Prerequisite explanations: in this method, the user model checks the prerequisites necessary to understand the content of the page.
- Comparative explanation : Comparative explanation is to explain new topics by stressing their relations to known topics
- Explanation variant: Explanation variants and extension to the prerequisite explanations
- Sorting: According to the need of the user, the different parts of the document are sorted.

Content level adaption methods will be implemented by the following techniques which deal with the knowledge. They are

- Conditional text: Information about a knowledge concept is divided in two different parts. Every part is defined with the knowledge.
- Stretch text: For some keywords of a document, according to the requirement of the user this technique provides longer descriptions.
- Page or page fragment variant: Different parts of the page are stored.
- Frame base fragments: this technique stores the page fragments into frames in a special order.

E. Link level adaption

Personalization for the user is being made through the link level adaption the following are the methods for navigating link level adaption[16].

i) Direct Guidance: “next best” and “page sequencing” are the two methods to guide the user sequentially through the hypermedia system [14]. “Nest best” provides nest button to navigate where page sequencing generates a reading sequence.

ii) Adaptive Sorting: “Similarity Sorting and “pre requisite sorting” are used based as the relevance system assumption by him/her, otherwise according to the prerequisite knowledge [8].

iii) Adaptive Hiding: Irrelevant information can be limited by making them unavailable or invisible.

iv) Link Annotation: Several methods are available to annotate the educational area links for example traffic metaphor, where a red ball indicates lack of knowledge of understanding the pages yellow ball indicates that the link to pages are not recommended for reading[7] [9]. Green ball indicates links recommended pages.

v) Map Annotation: the same link annotation methods can be applied for maps.

V. Adaptive Educational

A. HYPERMEDIA SYSTEM METHODOLOGY (AEHSM) :

A component based logical description of adaptive educational hypermedia system is proposed by Matteo et al [12]. This component based definition is based on the theory of diagnosis by Reiter [20].

B. How it works?

According to Matteo et al [12] AEHS was decomposed into basic components according to their roles. This uses a user model to model various characteristics of individual users or user groups. The adaptive functionality is provided by the organization of the document space and the user model [10].

This Adaptive Educational Hyper Media System is a quadruple. They are i) document Space (Docs), User Model (UM) Observations (OBS) and Adaption Component (AC) [22]. The document space and observations describe basic data and runtime data. This data will be processed by the other two. AEHS makes it Simple by annotating text using the traffic light metaphor. This can be extended by using Knowledge graph instead of domain graph [15]. This system is able to give a more differentiated traffic light annotation to hypertext links than simple [13]. It is able to recommend pages with green icon and to show which links lead to documents that will become understandable with dark orange icon and yellow icon is for the pages which might be understandable and red icon for which are not recommended yet. The representation of AEHS Simple and Knowledge graph with quadruple were presented in detail with examples.

a) Simple can annotate hypertext links by using the traffic light metaphor with two colors: red for non-recommended, green for recommended pages.

i) *DOCSs*: This component is made of a set of *n* constants and a finite set of predicates. Each of the constants represents a document in the document space (the documents are denoted by *D*1, *D*2, . . ., *Dn*). The predicates define pre-requisite conditions, i.e. they state which documents need to be studied before a document can be learned, e.g. *preq(Di,Dj*) for certain *Di _= Dj* means that *Dj* is a prerequisite for *Di*

ii) *UMs*: it contains a set of *m* constants, one for each individual user *U*1, *U*2, ..., *Um*.

iii) *OBSs*: A special constant (*Visited*) is used within the special predicate *obs* to denote whether a document has been visited: *obs* (*Di*, *Uj*, *Visited*) is the observation that a document *Di* has been visited by the user *Uj*.

iv) *ACs*: This component contains constants and rules. One constant is used for describing the values of the “learning state” of the adaptive functionality, two constants (*Green Icon* and *Red Icon*) for representing values of the adaptive functionality. The learning state of a document is described by a set of rules of kind:

$$\forall U_i \forall D_j(\forall D_k \text{preq}(D_j,D_k) \Rightarrow \text{obs}(D_k, U_i, \text{Visited})) \Rightarrow$$

$$\text{learning-state}(D_j, U_i,\text{Recommended for reading})$$

This component contains also a set of rules for describing the adaptive link annotation with traffic lights. Such rules are of kind:

$$\forall U_i \forall D_j \text{ learning- state}(D_j, U_i,\text{Recommended for- reading})$$

$$\Rightarrow \text{document annotation}(D_j, Ui,\text{Green\_icon})$$

or of kind:

$$\forall U_i \forall D_j \neg \text{ learning- state}(D_j, U_i,\text{Recommended for-reading})$$

$$\Rightarrow \text{document annotation } (D_j, U_i,\text{Green\_icon})$$

b) This simple AEHS can be extended by using a *knowledge graph* instead of a domain graph. The system, called Simple1, is able to give a more differentiated traffic light annotation to hypertext links than Simple [8]. It is able to recommend pages (green icon), to show which links lead to documents that will become understandable (dark orange icon), which might be understandable (yellow icon), or which are not recommended yet (red icon) [21]. Let us represent Simple1 by a quadruple (*DOCSs*1, *UMs*1, *OBSs*1, *ACs*1):

i) *DOCSs*1: The document space contains all axioms of the document space of Simple, *DOCSs*, but it does not contain any of the predicates. In addition, it contains a set of *s* constants which name the knowledge topics T1, T2, Ts in the knowledge space. It also contains a finite set of predicates, stating the learning dependencies between these topics: depends (Tj, Tk), with Tj _= Tk, means that topic Tk is required to understand Tj. The documents are characterized by predicate keyword which assigns a nonempty set of topics to each of them, so $\forall Di \exists$Tjkeyword (*Di*, *Tj)*, but keep in mind that more than one keyword might be assigned to a same document.

ii) *UMs*1: The user model is the same as in Simple, plus an additional rule which defines that a topic Ti is assumed to be learned whenever the corresponding document has been visited by the user. To this aim, Simple 1 uses the constant Learned. The rule for processing the observation that a topic has been learned by a user is as follows (p obs is the abbreviation for “processing an observation”):

$$\forall U_i \forall T_j (\exists D_k \text{keyword}(D_k, T_j) \wedge \ obs\ (D_k, U_i, \text{Visited})$$

$$\Rightarrow P\text{\_obs } (T_j, U_i,\text{Learned})$$

iii) OBSs1: Are the same as in Simple.

iv) ACs1: The adaptation component of Simple1 contains two further constants (w.r.t. Simple), representing new values for the learning state of a document [7] [4]. Such constants are: Might be understandable and will become understandable

Two more constants are added for representing new values for adaptive link annotation. They are: Orange Icon and Yellow Icon. Such constants appear in the rules that describe the educational state of a document, reported hereafter. The first rule states that a document is recommended for learning if all the prerequisites to the keywords of this document have already been learnt:

$$\forall U_i \forall D_j(\forall T_k \text{ keyword}(Dj, Tk) \Rightarrow$$

$$(\forall T_l \text{ depends}(T_k, T_l) =\Rightarrow \text{p\_obs}(T_l, U_i, \text{Learned})$$

$$\Rightarrow \text{learning\_state}(D_j, U_i,\text{Recommended\_for\_reading})))$$

The second rule states that a document might be understandable if at least some of the prerequisites have already been learnt by this user:

$$\forall U_i \forall D_j (\forall T_k \text{ keyword}(D_j, T_k) \Rightarrow$$

$$(\exists T_l \text{ depends}(T_k, T_l) \Rightarrow$$

$$\text{P\_obs } (T_l, U_i, \text{Learned})$$

$$\wedge \neg \text{learning state } (D_j, U_i, \text{Recommended\_for\_reading})$$

$$\Rightarrow \text{learning state } (D_j, U_i, \text{Might be understandable})))$$

The third rule entails that a document will become understandable if the user has some prerequisite knowledge for at least one of the document’s keywords:

$$\forall U_i \forall D_j (\exists T_k \text{ keyword } (D_j, T_k) \Rightarrow$$

$$(\exists T_l \text{ depends } (T_k, T_l) \Rightarrow$$

$$p\ obs(Tl, Ui, Learned)$$

$$\wedge \neg \text{learning state } (D_j, U_i, \text{Might be understandable})$$

$$\Rightarrow \text{Learning state}(D_j, U_i, \text{Will become understandable})))$$

Four rules describe the adaptive link annotation:

1) $U_i \ \forall D_j$ learning state(D_j, U_i,Recommended for reading)

$$\Rightarrow \text{document annotation } (D_j, U_i, \text{Green Icon})$$

2) $\forall \ U_i \ \forall \ D_j$ learning state (D_j, U_i,Will become \ understandable)

$$\Rightarrow \text{document annotation } (D_j, U_i,\text{Orange Icon})$$

3) $\forall U_i \forall D_j$ learning state(D_j, U_i, Might be understandable)

$\Rightarrow$ document annotation(D_j, U_i, Yellow Icon)

4) $\forall$ U_i $\forall$ D_j $\neg$ learning state(D_j, U_i,Recommended_for_reading)

$\Rightarrow$ document annotation(Dj, U_i,Red Icon)

VI. Conclusion

Present and the future research in e-learning system are on the intelligent learning systems. The platform for this is the Semantic Web and the Web Ontology's. One common assumption is that the Semantic Web can be made a reality by gradually augmenting the existing data (HTML/XHTML) by ontological annotations, derived from the on-machine-readable content This paper presents an intelligent e-learning system i.e., An Adaptive Educational Hyper media System which is based on the hypermedia system using hypertext link by traffic metaphor. This system is aimed at providing user required information effectively and efficiently. The aim of this study is to extend this model to other areas like e-commerce, Artificial intelligence problems for knowledge retrieval.

References

[1] Zhiming Cui, Wei Fang, Xuefeng Xian, Shukui Zhang, and Pengpeng Zhao "Extension of OWL with Dynamic Fuzzy Logic"proc. International workshop on web based contents managementtechnologies [WCMT-2009] pp 67-76

[2] Fayed F. M. Ghaleb, 1Sameh S. Daoud, 2Ahmad M. Hasna, 2Jihad M. Jaam and Hosam F. El-Sofany "A Web-Based E-Learning System Using Semantic Web Framework" Journal of Computer Science 2 (8): 619-626, 2006, ISSN 1549-3636, © 2006 Science Publications

[3] Nicola Capuano1, 3, Matteo Gaeta1, 3, Alessandro Micarelli1, 4 and Enver Sangineto2 "An Intelligent Web Teacher System for Learning Personalization and Semantic Web Compatibility".

[4] IWT: Intelligent Web Teacher. White Paper,CRMPA, 2002.

[5] Capuano N., De Santo M., Marsella M., Molinara M. and Salerno S. "Personalised Intelligent Training on the Web" Multimedia Systems and Applications Series, vol. 20, chap. 5, 2001.

[6] Capuano N., Gaeta M., Micarelli A. and Sangineto E. "An integrated Architecture for Automatic Course Generation". Proceedings of the IEEE International Conference on Advanced Learning Technologies, Kazan, Russia, 2002

[7] Felder R. M., "Learning and Teaching Styles in Engineering Education", Engr. Education 78 (7), 674-681, 1988.

[8] Matteo Baldoni1, Cristina Baroglio1, and Nicola Henze2" Personalization for the Semantic Web*" N. Eisinger and J. Maluszynski (Eds.): REWERSE 2005, LNCS 3564, pp. 173–212, 2005. c_Springer-Verlag Berlin Heidelberg 2005.

[9] http://swoogle.umbc.edu/

[10] Cardoso, J. and A. Sheth(2003). "Semantic e-Workflow Composion."Journal of Intelligent Information Systems (JIIS).21(3):191-225.

[11] Kashyap, V. and A. Sheth (1996). "Semantic heterogeneity in global information systems: the role of metadata, context and ontologies". Cooperative Information Systems: Current Trends and Applications. M.Papzoglou and G.Schlageter. London,Uk, Academic Press: 139-178.

[12] Matteo Baldoni, C. Baroglio, and V. Patti. "Web-based adaptive tutoring: an approach based on logic agents and reasoning about actions". Artificial Intelligence Review, 22(1), September 2004.

[13] Fensel, D., F. van Harmelen, I. Horrocks, D.L. Mc Guinness and P.F. Patel-Schneider, 2001. OIL: An ontology infrastructure for the semantic web. IEEE Intelligent Systems, 16: 38-45.

[14] Matteo Baldoni, C. Baroglio, V. Patti, and L. Torasso. "Reasoning about learning object metadata for adapting scorm courseware". In L. Aroyo and C. Tasso, editors, Proc. of Int. Workshop on Engineering the Adaptive Web, EAW'04: Methods and Technologies for personalization and Adaptation in the Semantic Web, pages 4–13, Eindhoven, The Netherlands, August 2004.

[15] Rokou, F.P. et al., 2004. "Modeling web-based educational systems: process design teaching model". Educat. Technol. Soc., 7: 42-50.

[16] Adelsberger, H. *et al*., 2003. "The Essen model: A step towards a standard learning process". http://citeseer.ist.psu.edu/515384.html.

[17] T. R. Gruber. "A translation approach to portable ontologies". Knowledge Acquisition, 5(2):199-220, 1993.

[18] T. R. Gruber. "Toward principles for the design of ontologies used for knowledge sharing". Presented at the Padua workshop on Formal Ontology, March 1993, later published in International Journal of Human-Computer Studies, Vol. 43, Issues 4-5, November 1995, pp. 907-928.

[19] Charlie Abela, Matthew Montebello "A Personalized Service Discovery and Composition Framework" funded by the European Commission and by the Swiss Federal Office for Education and Science within the 6th Framework Programme project REWERSE number 506779

[20] R. Reiter. "A theory of diagnosis from first principles" Artificial Intelligence, 32, 1987.

[21] J.Lobo, G.Mendez, and S.R. Taylor. "Adding Knowledge to the Action Description Language" A. In proc. Of AAAI97/IAAF97, pages 454-459 Menlo park,1997.

[22] I. Horrocks, P. Patel-Schneider, H. Boley. S. Tabet, and B. Grosof.SWRL: "A semantic web rule language combining OWL and RuleML",200

Studying Data Mining and Data Warehousing with Different E-Learning System

Dr. Mohamed F. AlAjmi,
PhD Head of Quality and
E-Leaning units
King Saud University
Riyadh, Saudi Arabia

Shakir Khan
M.Sc (Computer Science)
Researcher at King Saud
University, Riyadh Saudi Arabia
Nationality Indian

Dr. Arun Sharma
Head,Department of Computer
Science Krishna Institute of
Engineering and Technology,
Ghaziabad-201206, INDIA

***Abstract*----Data Mining and Data Warehousing are two most significant techniques for pattern detection and concentrated data management in present technology. ELearning is one of the most important applications of data mining. The foremost idea is to provide a proposal for a practical model and architecture. The standards and system structural design are analyzed here. This paper provides importance to the combination of Web Services on the e-Learning application domain, because Web Service is the most complex choice for distance education during these days. The process of e-Learning can be promising more efficiently by utilizing of Web usage mining. Mor07/e sophisticated tools are developed for internet customer's behaviour to boost sales and profit, but no such tools are developed to recognize learner's performance in e-Learning. In this paper, some data mining techniques are examined that could be used to improve web-based learning environments.**

Keywords— Data Mining; Data Warehousing; e-Learning; Moodle; LMS; LCMS.

I. Introduction

Usually the decision-making data are stored in files and databases. The results getting by huge amount of data are not easy, for which the data mining techniques are very constructive. Data mining is the process of taking out information in terms of patterns or set of laws (e.g. association rules, sequential patterns, classification trees) from huge databases. So, it is also known as data or knowledge discovery.

For example, by pulling out demographic data of students' enrolments, the university, college or any institute could get better the qualitative explanation (e.g. information for past's students) of database. Any association does not deal with a single database, but deals with various kind of database means multiple databases but there is the need for fast processing, and integrating of these databases which can be possible by data warehouse. Centralizing data management and revival is often distinct as data warehousing. This centralizing helps the user to maximize access to the data and analyzing it.

The data warehouse supports different types of analyses, including elaborate queries on large amounts of data that may require extensive searching. When databases are set up for queries on daily transactions, they are called "operational data stores" rather than data warehouse. So, a data warehouse is a storehouse of an organization's electronically stored data [3]. The mechanisms of data warehouse are: retrieval, extract, analysis, transform, load data and managing data dictionary. Data mining, data warehousing, and Online Analytical Processing (OLAP) together form the functionality of decision making or Decision Support System (DSS). The various areas Eof application of data mining and data warehousing are e-commerce, e governance, online shopping, digital library, online reading, e-learning or e-education, etc. Among these, these days e learning is an important application of data mining.

E-Learning is sometimes known as electronic learning or e-learning in which there is no face-to-face interaction between the teacher and the students. Rather than it is web-based learning. It uses Web or Internet technology and delivers digital contents, provides learner oriented environment for teachers and students [4]. So, the environment is not teacher-centric. It may include all types of Technology Enhanced Learning (TEL), where technology is used to support the learning process [5].

For example, in companies, e-Learning is used to deliver training courses to employees and in universities, e- Learning is used for enrolment of students in different courses, provides teaching without any face-to-face interaction, or on-campus facilities, but through internet that is online. As a whole, e-Learning includes Distance Learning (DL), Computer Based Teaching (CBT), Computer Aided Instruction (CAI), and Life Long Learning (LLL) principle. So, we see that, e-Learning consists of various types of databases, storing information for user access. To implement e-Learning, data mining can help to construct e-textbook, e-reading, digital libraries, etc.

Further scope of e-Learning is blended e- Learning which is a combination of face-to-face interaction and online learning. It incorporates online lectures, tutorials, performance and decision support systems, simulations and games, and more [5].

II. E-Learning Architecture or Design

A. *Functional Model*

The practical model of an e-Learning structure creates an interface between the mechanisms and the objects of the e-Learning system. It is shown in "Fig. 1".

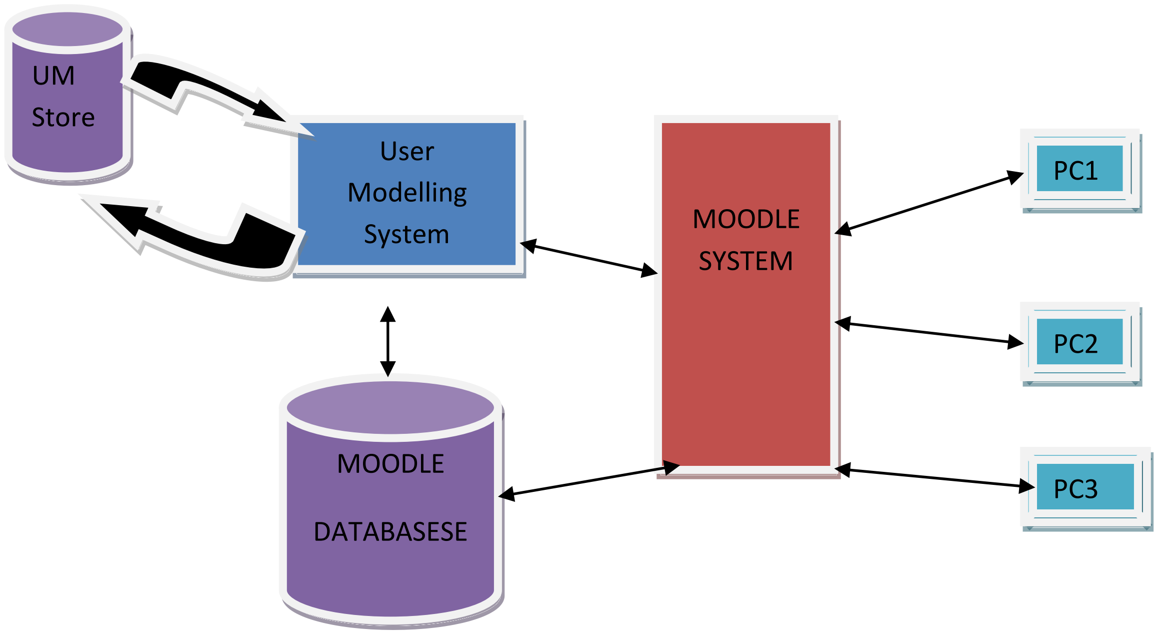


Fig. 1 Moodle Based E-Learning Architecture

The structural design of e-Learning till now does not provide any apparent picture of the e-Learning components. The e-Learning structural design contains two models: the information model and the component model. These two replicas are to be joined and an interface must be defined to attain interoperability. This structural design of e-Learning gives a practical model of the components of e-Learning for the consistency of e-Learning development. The Advanced Distributed Learning (ADL)'s Sharable Content Object Reference Model (SCORM) practical model explains the swap of data within a Learning Content Management System (LCMS) or a Learning Management System (LMS) to track user's progress. But the functionality is not explained by SCORM.

A multi-user atmosphere in which the knowledge developer can create, reuse, manage, store, and distribute digital learning content from a central storehouse is known as LCMS. Here the processes adjoining the learning are managed by LMS. LCMS permits the users to generate and to use again small units of digital instructional learning material.

The incorporated use of metadata arrangements and learning object import and export formats also allows learning objects to be created and shared by multiple tools and repositories. LCMS integrates specifications of metadata, content wrapping, and content communication. The components of LCMS are shown in "Fig 1".

LMS needs the interchange of customer profile and customer registration information with other systems. The position of the course choice and the learner action are offered by the LCMS. The mechanisms and information needed are shown in "fig 1". So, there is an incorporation of LMS and LCMS.

Secondly, the SCORM is developed by US Department of Defense's ADL. This is an "application profile" consisting of a set of terms and conditions. The three main mechanisms of SCORM are:

1) *Runtime Environment: The runtime environment is an API describes the interface between learning object and LMS or LCMS to track learner's progress;*
2) *Meta-Data: A set of data elements to explain learning contents so that it can simply explore for identified and accessed [7];*
3) *Content wrapping: Content wrapping is the release and exchange of structured content i.e. learning objects and courses between different LMS and LCMSs;*

As a course is separated into lessons, and sometimes the lessons are divided into topics The SCORM condition explains two hierarchical levels:

1) *Content aggregation: A group of learning resources to construct complex structures, contents aggregations may be nested and may have lower-level blocks of contents which outline a content aggregation;*
2) *Resources: Two major types of educating resources are there: SCO and ASSET;*

The stage at which student interacts with the learning content and also the LMS tracks the results is known as SCO. Basically, it is a learning object.

A part of content in form of movie, sound, graphic or other media item is referred as an ASSET. Most ASSETS are started by SCOs as part of their in-house content (e.g. graphics come into view on an HTML page).

III. Standards In E-Learning

Standards in e-Learning give standardized data structures and communication protocols for e-Learning objects and cross-system workflows [1]. The standards are of the following types:

1) *Metadata: Metadata refers to the labelling of learning contents and catalogs to maintain indexing, storage, detection (searching), recovery of learning objects by several repositories of data mining and data warehousing techniques. The data utilized here is known as metadata;*
2) *Content Wrapping: Content Wrapping permits the transport of course content from one learning management system to another learning management system. The most significant content wrapping system these days is, ADL's SCORM [7]. The facts of the contents are stored in various databases which can be developed and received by data mining and data warehousing techniques;*
3) *User Profile: User Profile consist personal data, learning history, prerequisites, learning plans, degrees and certifications, evaluation of information and contribution status in existing learning;*
4) *Student Registration: Student Registration identifies the availability of courses for the learner, also, information about other members of the course.*
5) *Content Communication: It gives an interface between student data and previous activity after content is started. The message is developed by ADL's SCORM Object Reference Model.*

This architecture explains the fundamental thought of scattered e-Learning system means the communication of messages through the communication of web service agents, present in each system. Service Provider is the podium that hosts right to use to the service. It is the server in a client-service environment. Service Requester is the function that is looking for and calling upon or initiating the communication with a service. Discovery Agency is a searchable set of service explanation where service providers issue their service descriptions.

According to Xiaofei Liu, Abdulmotaleb EI Saddik and Nicolas D. Georganas [1], the discovery agency may be centralized or distributed. Information presented by XML concerning learning is wrapped with the Simple Object Access Protocol (SOAP) arrangement and is swapped between requester and provider. A Web Services Description Language (WSDL) file holding the explanation of the message and information regarding end point is published by the provider to permit requester to create the SOAP message and transmit it to the exact destination.

IV. Benefits Of Data Mining In E-Learning

There are several web usage tools to carry out data mining and data ware housing tasks. For, instance, Two data mining and data ware housing tools are WebSIFT and WebLogMiner for pattern detection from web logs [10][11] but these tools are not initiated in e-Learning environment till now because if the educator does not have sufficient knowledge in data mining, can't use these tools to get better efficiency of e-Learning. Web usage mining is a new system, devoted for e-Learning is being industrialized to permit the educators for on-line assess activities [9]. It facilitates the educator to follow the activities in the course web site and take out patterns and behaviours, get better or adapt the course content. For example, one could recognize the paths regularly or frequently visited, the paths never visited, etc. By analyzing these general traversal paths of the course content web pages or recurrent changes in individual traversal paths, the design of the course can be known to be better fit the requirements of students.

Two types of data mining techniques are used in e-Learning: off-line web usage mining and integrated web usage mining. Off-line Web Usage Mining: Off-line web usage mining is the detection of patterns with a separate application. This pattern detection process permits educators to evaluate the access behaviours, legalization of the learning modules, assessment of the learner's activities, assessment between learners and their access pattern, etc [9]. The model of off-line web mining is a tool for the instructors to apply sequential analysis, association rules and clustering for the detection of relations between the learning actions of learners, interesting prototype of on-line actions and to group parallel access behaviour respectively. So, in off-line web usage mining, incorporated educators can place questions and authenticate the learning models, they utilize as well as the structure of the web site as it is read thoroughly by the learner. It is being observed that off-line web usage mining is a parametric move towards where the parameters are the instructors, educators, learners, etc.

Integrated Web Usage Mining: Contrasting to off-line web usage mining, incorporated web usage mining is the procedure of determining patterns incorporated with e-Learning application. This covers adaptive websites, personalization of actions. Also, suggestion of actions to learners according to their favourites along with their history of actions is done by automatic recommenders in incorporated web usage mining. A recommender-based association rule mining is being expanded currently that consists of facts of finding out applicable association between learning performance and creating association rules are recommended to the learner as the suggested next step in the learning session [10]. So, incorporated web usage mining is a non-parametric approach.

V. Conclusion And Future Work

In this paper, an obvious analysis of the content state of e-Learning standard is being explained. Also, a functional model of dissimilar learning objects is presented here. The swapping of system workflows is also being explained in this paper. E-learning standard gives interoperability between learning systems and tools from several vendors. A standard means of message is set up between dissimilar software applications. This communication is likely by the Web-Services technology.

The Web usage mining technique is explained in this paper, which is a non-trivial procedure of taking out helpful and previously unknown blueprints from the use of Web. The data mining techniques to improve e-education are explained in this paper. Since e-Learning process is a endlessly changeable process, the safety services, the encryption of messages, and the general facts to explain services and services access points in e-Learning systems environments are in call for thought.

Though, several tools using data mining techniques to aid e-Learning system are being developed, the research is still in progress, since the data record given by the Web Servers are inadequate, so there is a call for more specialized logs from the application side to improve the already logged information.

REFERENCES

[1] Xiaofei Liu, Abdulmotaleb El Saddik and Nicolas D.Georganas"AN IMPLEMENTABLE ARCHITECTURE OF AN E-LEARNING SYSTEMS". CCECE 2003 – CCGEI 2003 Montreal May/mai 2003.

[2] FUNDAMENTALS OF DATABASE SYSTEMS, Fourth Edition, Elmasri and Navathe.

[3] Data Mining: What is Data Mining, Web site at http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm.

[4] "Introduction to E-Learning "Website at http://www.chengzhi.net/english/index.htm

[5] "E-Learning From Wikipedia to free encyclopedia", Web site at http://en.wikipedia.org/wiki/E-Learning.

[6] C. Romero and S. Ventura , "Data Mining in E-Learning" (WIT Press ,2006).

[7] Leopold Kause, Carol Fallon "Creating E-Learning Content in Authorware 7 for SCORM1.2 Compliant LMSs and LCMSs", Web site at http://adobe.com/resources/elearning/article/l0packager01.

[8] "IMS Global Learning Consortium", Web site at http://www.imsproject.org/

[9] Osmar R. Za¨ıane, "Web Usage Mining for a Better Web-Based Learning Environment".

[10] R. Cooley, B.Mobasher, J.Srivastva, "Web Mining: Information and Pattern Discovery on the World Wide Web", Procedings of the ninth IEEE international conference on Tools with AI, 1997.

[11] O.R. Zaiane, M.Xin, J. Han, Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs, Proceedings from the ADL'98 –Advances in Digital Libraries, Santa Barbara, 1998.

[12] E-Learning: A Milestone in the Research of Data Mining; by Sabyasachi Pattnaik, Jui Pattnayak, Priyaranjan Dash http://interscience.in/SpIss_ijcct_icct2010vol2_no234/25_ML

AUTHORS PROFILE

Dr Mohammed Fahad AlAjmi was born in Kingdom of Saudi Arabia. He received his Ph.D in Pharmacy from King Saud University, Riyadh, Saudi Arabia in 2007.He chaired many position in the university and currently working as vice dean for quality and development in Prince Sultan College for EMS affiliated to King Saud University. To date he taught many pharmacy students, more than 30 courses. Students' level varies from primary to undergraduate levels.

Shakir Khan was born on 5th Feb, 1978 at Kallanheri in Saharanpur district UP, India. He is working as a Researcher in College of Electronic Learning in King Saud University, Kingdom of Saudi Arabia. He received his Master of Science in Computer Science from Jamia Hamdard (Hamdard University), New Delhi, India in the year 2005, and PhD computer Science scholar in Manav Bharti University, Solan (HP) India since 2010. He is member of IEEE. He has actively attended many international conferences and published various research papers in National and International conferences as well as journals. His current areas of interests are in Cloud Computing, Software Engineering, Data Mining and E Learning. Apart from that he worked in the field of Software Development in different MNC companies at Noida India

Dr. Arun Sharma, alumni of IIT Roorkee and Thapar University, received his M.Tech. (Computer Science and Engineering) from Punjabi University, Patiala, INDIA and Ph.D. (Computer Science) Thapar University, Patiala, INDIA. Currently, he is working as head of the department of Computer Science and Engineering Department in KIET school of engineering and Technology at Ghaziabad, India. His areas of interests include Software Engineering, Soft Computing and Database Systems. He has published a number of papers in international Journals and Conferences including IEEE, ACM, Springer, WILEY and others in India and abroad. Dr. Sharma is an active member of IEEE, ACM and Computer Society of India. He is also a member of Board of Studies (BoS) of Mahamaya Technical University (MTU), Noida. He is also on the panel of subject experts and examination for various Universities like IGNOU, BBA University (Central), Lucknow, GGSIP University, Delhi, Thapar University, Patiala, and others. He is also an active member of Editorial Board and Review Committee of several Journals including Journal of Computer Science (USA), International Journal of Computer Science and Security (Malaysia), Research Journal of Information Technology, USA and others.

Monte Carlo Ray Tracing Based Non-Linear Mixture Model of Mixed Pixels in Earth Observation Satellite Imagery Data

Verification of non-linear mixed pixel model with real remote sensing satellite images

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

***Abstract*—Monte Carlo based non-linear mixel (mixed pixel) model of visible to near infrared radiometer of earth observation satellite imagery is proposed. Through comparative studies with actual real earth observation satellite imagery data between conventional linear mixel model and the proposed non-linear mixel model, it is found that the proposed mixel model represents the pixels in concern much precisely rather than the conventional linear mixel model.**

Keywords-remote sensing satellite; visible to near infrared radiometer; mixed pixel: mixel; Monte Carlo simulation model

I. Introduction

The pixels in earth observed images which are acquired with Visible to Near Infrared: VNIR sensors onboard remote sensing satellites are, essentially mixed pixels (mixels) which consists of several ground cover materials [1]. Some mixel model is required for analysis such as un-mixing of the mixel in concern [2],[3]. Typical mixel is linear mixing model which is represented by linear combination of several ground cover materials with mixing ratio for each material [4]. It is not always true that the linear mixel model is appropriate [5]. Due to the influences from multiple reflections between the atmosphere and ground, multiple scattering in the atmosphere on the observed radiance from the ground surface, pixel mixture model is essentially non-linear rather than linear. These influence is interpreted as adjacency effect [6], [7].

Method for representation of non-linear mixel model is not so easy. In particular, there is not sophisticated multi reflection model between ground materials. The representation method for non-linear mixel model is based on Monte Carlo Ray Tracing: MCRT model [8]. It is rather easy to designate surface slopes on the ground and multiple reflection among trees for MCRT model. The proposed MCRT based non-linear mixel model is applied to real earth observation satellite imagery data of Advanced Spaceborn Thermal Emission and Reflection Radiometer / Visible and Near Infrared Radiometer: ASTER/VNIR onboard on Terra satellite. A comparison of radiance between the conventional linear mixel model and the proposed non-linear mixel model is conducted. As a result, validity of the proposed model is confirmed.

The following section describes the proposed non-linear mixel model based on MCRT followed by some experiments for validation of the proposed model. Then, finally, conclusions with some discussions are described.

II. Proposed Non-Linear Mixel Model

A. Monte CarloRay Tracing Simulation

In order to show a validity of the proposed non-linear mixel model, MCRT simulation study and field experimental study is conducted. MCRT allows simulation of polarization characteristics of sea surface with designated parameters of the atmospheric conditions and sea surface and sea water conditions. Illustrative view of MCRT is shown in Figure 1.

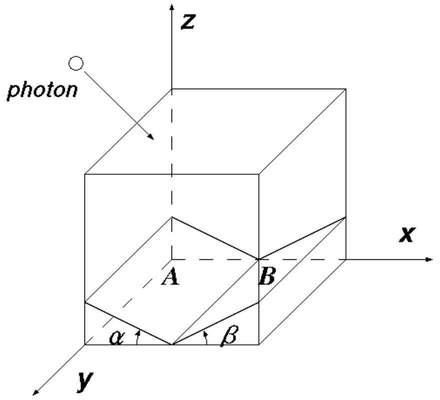
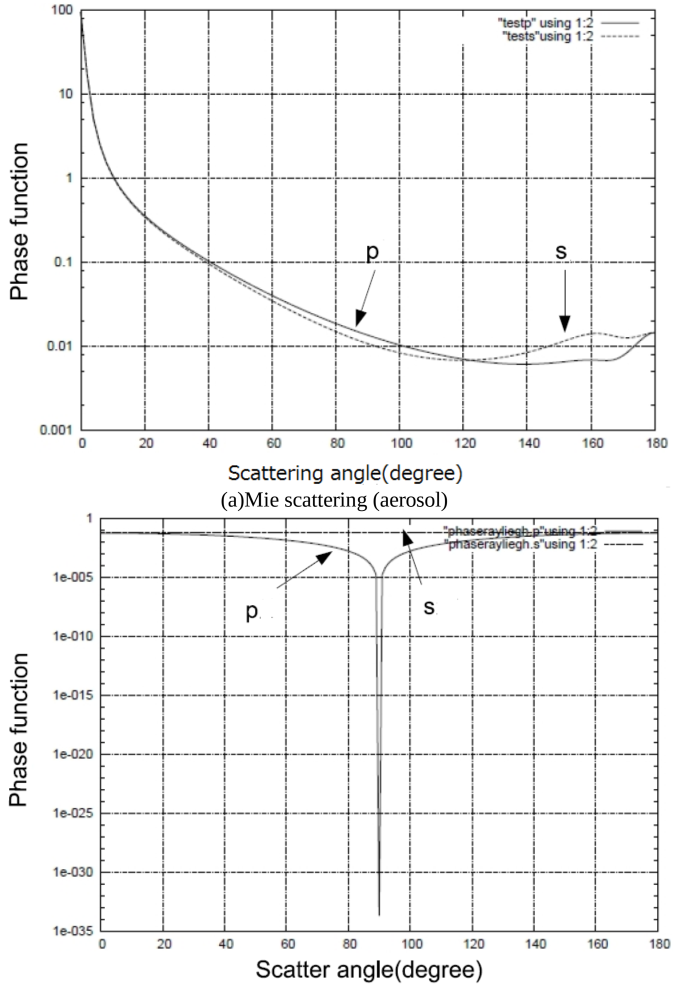


Figure 1 Illustrative view of MCRT for the atmosphere and sea water

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.

The scattering property is called as phase function[1]. In the visible to near infrared wavelength region, the scattering by molecule is followed by Rayleigh scattering law [10] while that by aerosol is followed by Mie scattering law [10]. Example of phase function of Mie scattering is shown in Figure 2 (a) while that of Rayleigh scattering is shown in Figure 2 (b).



(a)Mie scattering (aerosol)

(b)Rayleigh scattering (Molecule)

Figure 2 Phase functions for Mie and Rayleigh scattering

In the atmosphere, there are absorption due to water vapor, ozone and aerosols together with scattering due to the atmospheric molecules, aerosols. Atmospheric Optical Depth: AOD (optical thickness) in total, Optical Depth: OD due to water vapor (H_2O), ozone (O_3), molecules (MOL), aerosols (AER), and real observed OD (OBS) are plotted in Figure 3 as an example.

For simplifying the calculations of the atmospheric influences, it is assumed that the atmosphere containing only molecules and aerosols. As shown in Figure 3, this assumption is not so bad. Thus the travel length of the photon at once, L is expressed with equation (1).

$$L=L_0\,\mathrm{RND}(i) \qquad (1)$$

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

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 aforementioned phase function and with the rest of the travel length of the photon.

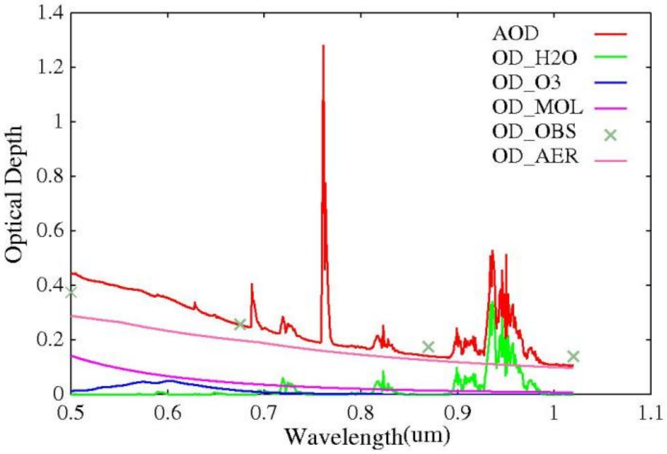


Figure 3 Example of observed atmospheric optical depth in total and the best fit curves of optical depth due to water vapor, ozone, molecules, and aerosols calculated with MODTRAN of atmospheric radiative transfer software code..

B. *Ground Surface with Slopes*

When the photon reaches on the ground, the photon reflects at the ground surface to the direction which is determined by random number. Lambertian surface [11] is assumed. Therefore, reflectance is constant for all the directions. The reflected photon travels with the rest of travel length. Two adjacent slopes of Lambertian surfaces are assumed on the ground as shown in Figure 4. Slope angles for both are α、β while their reflectance are Γ_A and Γ_B

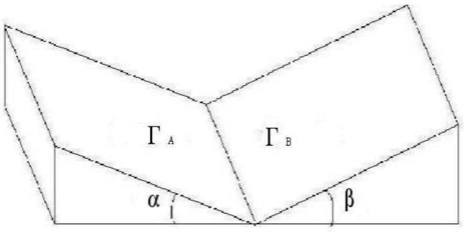


Figure 4 Two adjacent slopes of Lambertian surfaces which are assumed on the ground

C. *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

[1] http://ejje.weblio.jp/content/phase+function

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 (3)

$$I^+(\mu, \mu_0)=I\, N^+(\mu, \mu_0)/N_{total} \qquad (3)$$

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.

III. Experiments

A. Validity of the Monte CarloRay Tracing Simulation

In order to confirm that the developed MCRT is valid, a comparative study is conducted between radiative transfer code of Gauss Seidel method and the MCRT derived TOA radiance. Because the Gauss Seidel method allows calculation of TOA radiance with flat surface of ground, 0.2 of reflectance of flat surface is assumed in the comparison. Also, 0.02 and 0.03 of optical depths are assumed for aerosol and molecule. The size of simulation cell is determined as 50 km by 50 km by 50 km. Solar zenith angle is set at 30 degree while solar azimuth is set at 120 degree. 700,000 of photons are input to the simulation cell. TOA radiance derived from the Gauss Seidel method is 0.565 (mW/m^2/sr/μm) while that from the MCRT is 0.579 (mW/m^2/sr/μm) at the 500nm of wavelength. For both cases, IFOV of the VNIR radiometer is assumed to be 2 π ; all of the photons output from the top of the atmosphere are counted. Therefore, the developed MCRT seems valid enough.

B. TOA Radiance for the Different Combination of Optical Depths of Aerosol and Molecule and for the Ground with the Different Slopes

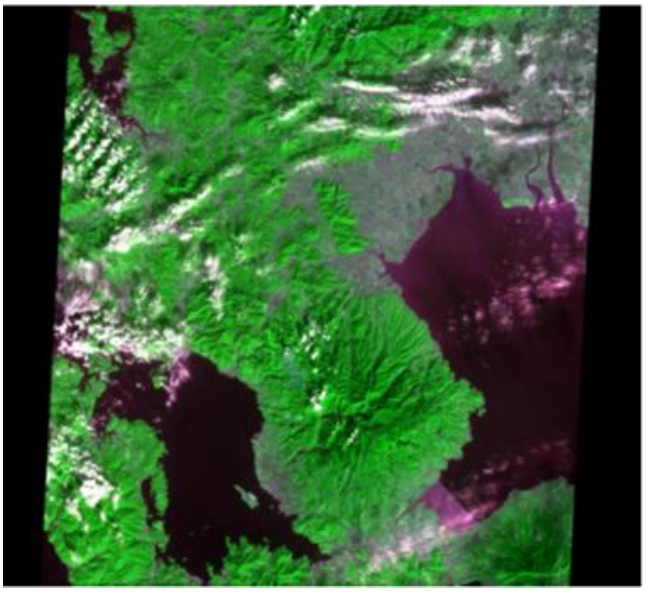
TOA radiance at 500 nm of wavelength for the different combination of optical depths of aerosol and molecule which ranges from 0.01 to 0.04 and for the ground with the different slopes, 0 and 20 degree are calculated. Again, IFOV of the VNIR radiometer is assumed to be 2 π, all of the photons output from the top of the atmosphere are counted. The reflectance for both slopes are same as 0.5. The results are shown in Table 1. In the table, τ_{aer} , τ_{mol} are optical depths of aerosol and molecule, respectively.

C. Validity of the Proposed Non-Linear Mixel Model with Real VNIR Data

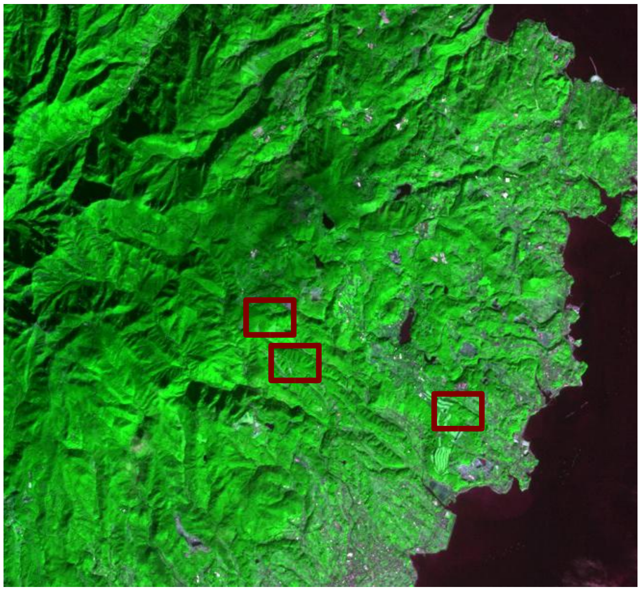
The proposed non-linear mixel model based on MCRT is validated with real earth observation satellite imagery data of ASTER/VNIR onboard Terra satellite [9] which is acquired at 11:09 Japanese Standard Time: JST on December 15 2004. IFOV of ASTER/VNIR is 15m with 60km of swath width. Whole scene of ASTER/VNIR is shown in Figure 5 (a) while Figure 5 (b) shows a portion of the scene.

TABLE I. Top of the Atmosphere: TOA Radiance for the Combinations of Atmospheric Conditions

	TOA radiance (mW/m^2/sr/μm)			
τ_{aer} \ τ_{mol}	0.01	0.02	0.03	0.04
0.01	3.51	0.317	3.99	6.51
0.02	5.65	1.91	1.09	3.04
0.03	5.7	3.08	0.622	10.7
0.04	3.29	3.85	3.97	7.45



(a)Whole scene of ASTER/VNIR image



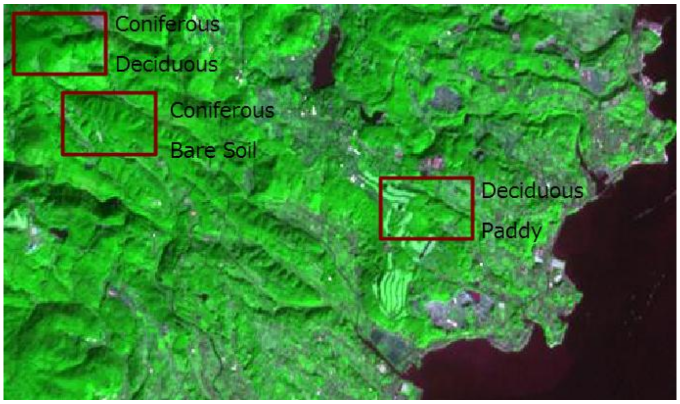
(b)A portion of the scene

Figure 5 ASTER/VNIR image used for experiment

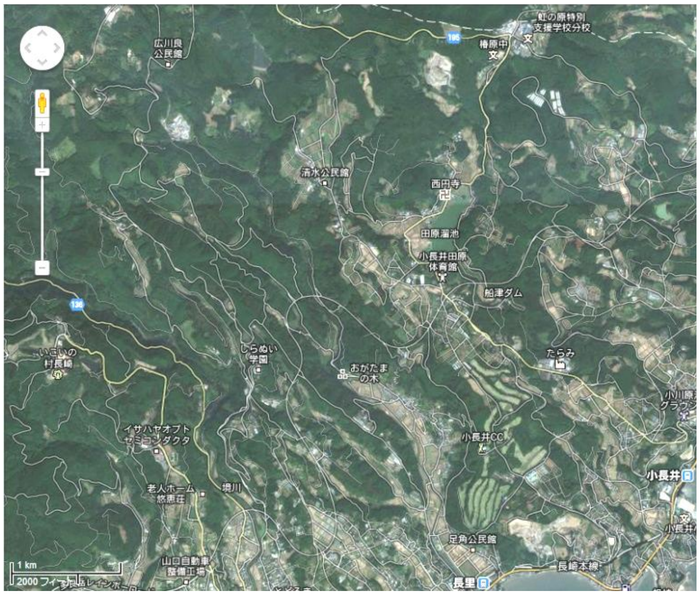
Three test sites, Area #1, 2, 3 are extracted from the scene. Attribute information of these sites are listed in Table 2.

TABLE II. ATTRIBUTIONS FOR THE TEST SITE WITH SLOPES

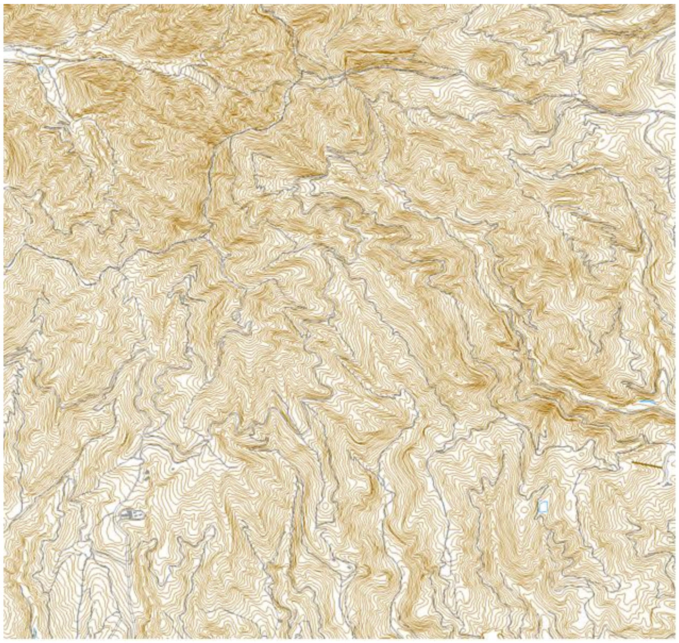
	Area #1	Area #2	Area #3
Area Name	Korai-cho, Ochiai-gawa	Korai-cho, Ochiai-gawa	Konagai Golf Club
Latitude	32°57'30"	32°56'33"	32°56'13"
Longitude	130°7'19"	130°7'25"	130°10'21"
Slope A(°)	24	30	20
Slope B(°)	28	26	0
Γ_A	0.14	0.2	0.14
Material	Deciduous	Bare Soil	Deciduous
Γ_B	0.08	0.08	0.12
Material	Coniferous	Coniferous	Paddy
OD-Aerosol	0.35	0.35	0.35
OD-Molecule	0.14	0.14	0.14



(a)Three test sites on ASTER/VNIR image



(b)Three test sites on Google Map



(c)topographic map of corresponding area of three test sites on Google map

Coniferous (above), Deciduous (bottom)

(d)Area#1 (Korai-cho, Ochiai-gawa, Nagasaki, Japan)

Coniferous (above), Bare Soil (bottom)

(e)Area #2 (Korai-cho, Ochiai-gawa, Nagasaki, Japan)

Deciduous (above), Paddy field (bottom)

(f)Area #3 (Konagai Country Club, Nagasaki, Japan)

Figure 6 Three test site, Area #1, 2, 3.

Figure 6 (a) shows three test sites on ASTER/VNIR image while Figure 6 (b) shows three test sites on Google map. Other than these, topographic map of three test sites which is corresponding to the Google map is shown in Figure 6 (c) while the extracted portion of each test site on ASTER/VNIR image is shown in Figure 6 (d), (e) and (f), respectively. These digital elevation models for three test sites are taken into account in the MCRT simulations. Also, solar zenith angle of 58 degree and solar azimuth angle of 17 degree are taken into account in the simulations. From the atmospheric optical depth measurement data with sun photometer, optical depth of total atmosphere is calculated. Furthermore, molecule optical depth τ_R is calculated with equation (4) as a function of atmospheric pressure P which is measured on the ground.

$$\tau_R(\lambda) = \frac{P}{P_0} \cdot 0.00864\lambda^{-(3.916+0.074\lambda+\frac{0.05}{\lambda})} \quad (4)$$

where P_0 denotes standard atmospheric pressure on the ground (1013 hPa) while λ denotes wavelength. Then aerosol optical depth is calculated from total atmospheric optical depth by subtracting molecule optical depth.

Comparative study is conducted between ASTER/VNIR derived radiance of Band 2 (Green band) and the radiance which derived from the conventional linear mixel model and the proposed non-linear mixel model. Table 3 shows the calculated radiance in unit of μW/m^2/sr/μm and the radiance difference between ASTER/VNIR and the estimated with the conventional and the proposed mixel models.

TABLE III. COMPARISON OF RADIANCE BETWEEN REAL ASTER/VNIR AND THE CONVENTIONAL LINEAR MIXEL MODEL AS WELL AS THE PROPOSED NON-LINEAR MIXEL MODEL DERIVED RADIANCE

	Area #1	Area #2	Area #3
ASTER/VNIR	14.1	15.5	16
Linear	12.9	13.7	14.6
Non-Linear	13.7	14.3	15
VNIR-Linear	1.2	1.8	1.4
VNIR-Non-Linear	0.4	1.2	1

It is found that the estimated radiance with the proposed non-linear mixel model is much closer rather than that with the conventional linear mixel model.

IV. CONCLUSION

Monte Carlo based non-linear mixel (mixed pixel) model of visible to near infrared radiometer of earth observation satellite imagery is proposed. Through comparative studies between ASTER/VNIR derived radiance and the conventional linear mixel model derived radiance as well as the proposed non-linear mixel model derived radiance, it is found that the estimated radiance with the proposed non-linear mixel model is much closer to ASTER/VNIR derived radiance (around 6%) rather than that with the conventional linear mixel model. One of the disadvantages of the proposed non-linear mixel model based on MCRT is time consumable computations. Acceleration is highly required.

ACKNOWLEDGMENT

The author would like to thank Dr. Yasunori Terayama and Mr. Kohei Imaoka of Saga University for their effort to simulation study and experiments.

REFERENCES

[1] Masao Matsumoto, Hiroki Fujiku, Kiyoshi Tsuchiya, Kohei Arai, Category decomposition in the maximum likelihood classification, Journal of Japan Society of Phtogrammetro and Remote Sensing, 30, 2, 25-34, 1991.

[2] Masao Moriyama, Yasunori Terayama, Kohei Arai, Clafficication method based on the mixing ratio by means of category decomposition, Journal of Remote Sensing Society of Japan, 13, 3, 23-32, 1993.

[3] Kohei Arai and H.Chen, Unmixing method for hyperspectral data based on subspace method with learning process, Techninical Notes of the Science and Engineering Faculty of Saga University,, 35, 1, 41-46, 2006.

[4] Kohei Arai and Y.Terayama, Label Relaxation Using a Linear Mixture Model, International Journal of Remote Sensing, 13, 16, 3217-3227, 1992.

[5] Kohei Arai, Yasunori Terayama, Yoko Ueda, Masao Moriyama, Cloud coverage ratio estimations within a pixel by means of category decomposition, Journal of Japan Society of Phtogrammetro and Remote Sensing, 31, 5, 4-10, 1992.

[6] Kohei Arai, Non-linear mixture model of mixed pixels in remote sensing satellite images based on Monte Carlo simulation, Advances in Space Research, 41, 11, 1715-1723, 2008.

[7] Kohei Arai, Kakei Chen, Category decomposition of hyper spectral data analysis based on sub-space method with learning processes, Journal of Japan Society of Phtogrammetro and Remote Sensing, 45, 5, 23-31, 2006.

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

[9] Ramachandran, Justice, Abrams(Edt.),Kohei Arai et al., Land Remote Sensing and Global Environmental Changes, Part-II, Sec.5: ASTER VNIR and SWIR Radiometric Calibration and Atmospheric Correction, 83-116, Springer 2010.

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

[11] 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.

Comparison and Analysis of Different Software Cost Estimation Methods

Sweta Kumari
Computer Science & Engineering
Birla Institute of Technology
Ranchi India

Shashank Pushkar
Computer Science &Engineering
Birla Institute of Technology
Ranchi India

Abstract- **Software cost estimation is the process of predicting the effort required to develop a software system. The basic input for the software cost estimation is coding size and set of cost drivers, the output is Effort in terms of Person-Months (PM's). Here, the use of support vector regression (SVR) has been proposed for the estimation of software project effort. We have used the COCOMO dataset and our results are compared to Intermediate COCOMO as well as to MOPSO model results for this dataset. It has been observed from the simulation that SVR outperforms other estimating techniques. This paper provides a comparative study on support vector regression (SVR), Intermediate COCOMO and Multiple Objective Particle Swarm Optimization (MOPSO) model for estimation of software project effort.**

We have analyzed in terms of accuracy and Error rate. Here, data mining tool Weka is used for simulation.

Keywords--- Support vector regression; PM- person-months; MOPSO- Multiple objective particle swarm optimization; COCOMO- Constructive cost estimation; Weka data mining tools.

I. INTRODUCTION

Cost estimation is a process or an approximation of the probable cost of a product, program, or a project, computed on the basis of available information. Accurate cost estimation is very important for every kind of project, if we do not estimate the projects in a proper way; result the cost of the project is very high sometimes it will be reached 150-200% more than the original cost [19]. So in that case it is very necessary to estimate the project correctly. The Cost for a project is a function of many parameters. Size is a primary cost factor in most models and can be measured using lines of code (LOC) or thousands of delivered lines of code (KDLOC) or function points. A number of models have been evolved to establish the relation between size and effort for Software Cost Estimation. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Data mining help us to classify the past project data and generate the valuable information.

Support vector regression (SVR) is a kernel method for regression based on the principle of structural risk minimization [11, 3]. Kernel methods have outperformed more traditional techniques in a number of problems, including classification and regression [11, 3]. Here, the use of SVR has been proposed for the estimation of software project cost and also, it has been found that this technique outperforms the other popular cost estimation procedures in terms of accuracy. The rest of the paper is organized as follows: Literature review refers to some existing estimation methods. Then the basic idea for this new approach for estimation has been discussed. Then the simulated experiment has been mention. We discuss the results and give the concluding remarks.

II. LITERATURE REVIEW

Various effort estimation models have been developed over the last four decades. The most commonly used methods for predicting software development efforts are function Point Analysis and Constructive Cost Model (COCOMO) [10]. Function point analysis is a method of quantifying the size and complexity of a software system in terms of the functions that the system delivers to the user [4]. The function does not depend on the programming languages or tools used to develop a software project [3]. COCOMO is developed by the Boehm [2]. It is based on linear-least-squares regression. Using line of code (LOC) as the unit of measure for software size itself contains so many problems [7]. These methods failed to deal with the implicit non-linearity and interactions between the characteristics of the project and effort [5, 11].

In recent years, a number of alternative modelling techniques have been proposed. They include artificial neural networks, analogy-based reasoning, and fuzzy system and ensemble techniques. Ensemble is used to combine the result of individual methods [12, 17]. In analogy-based cost estimation, similarity measures between a pair of projects play a critical role [16]. This type of model calculates distance between the software project being estimated and each of the historical software projects and then retrieves the most similar project for generating an effort estimate [14]. Further, Lefley and Shepperd [9] applied genetic programming to improve software cost estimation on public datasets with great success. Later, Vinay kumar et al. [15] used wavelet neural networks for the prediction of software cost estimation. Unfortunately the accuracy of these models is not satisfactory so there is always a scope for more accurate software cost estimation techniques.

III. THE BASIC IDEA

Suppose we are given training dataset$\{(x_1, y_1), \ldots, (x_l, y_l)\} \subset \chi \times \mathbb{R}$, where χ denotes the space of the input patterns (e.g. $\chi = \mathbb{R}^d$). The goal of regression is to find the function $f(x)$ that best models the training data. In our case, we are interested in building a regression model based on the training data to use it subsequently to predict the total effort in man-months of future software projects. In linear regression, this is done by finding the line that minimizes the sum of squares error on the training set.

A. Support Vector Regression

In this work we propose to use ε-SVR, which defines the ε-insensitive loss function. This type of loss function defines a band around the true outputs sometimes referred to as a tube, as shown in Fig. 1.

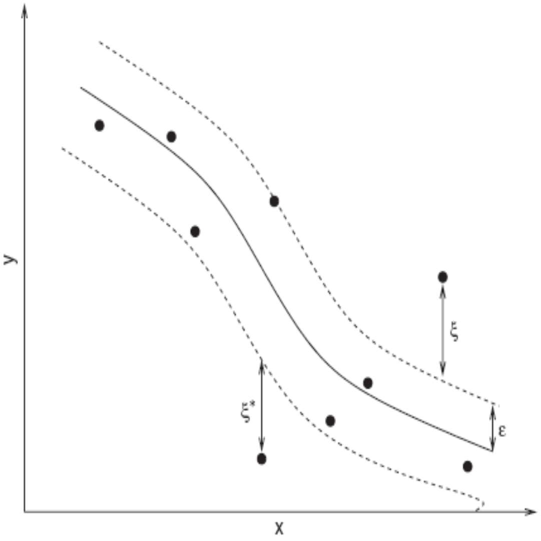


Fig.1 Regression using ε-SVR

The idea is that errors smaller than a certain threshold $\varepsilon > 0$ are ignored. That is, errors inside the band are considered to be zero. On the other hand, errors caused by points outside the band are measured by variables ξ and ξ* as shown in Fig. 1.

In the case of SVR for linear regression, $f(x)$ is given $f(x) = \langle w, x \rangle + b$, with $w \in \chi$, $b \in \mathbb{R}$. $\langle .,. \rangle$ denotes the dot product. For the case of nonlinear regression, $f(x) = \langle w, \phi(x) \rangle + b$, where ϕ is some nonlinear function which maps the input space to a higher (maybe infinite) dimensional feature space. In ε-SVR, the weight vector *w* and the threshold *b* are chosen to optimize the following problem [11]:

$$\text{minimize}_{w,b,\xi,\xi^*} \ \frac{1}{2} \langle w, w \rangle + C \sum_{i=1}^{l} (\xi i + \xi i^*),$$

$$\text{subject to} \begin{cases} (\langle w, \phi(\mathcal{X}_i) \rangle + b) - y_i \leq \varepsilon + \xi i, \\ y_i - (\langle w, \phi(\mathcal{X}_i) \rangle + b) \leq \varepsilon + \xi i^*, \\ \xi i, \ \xi i^* \geq 0 \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots (1) \end{cases}$$

The constant C>0 determines the trade-off between the flatness of *f* and the amount up to which deviations larger than ε are tolerated. ξ and ξ* are called slack variables and measure the cost of the errors on the training points. ξ measures deviations exceeding the target value by more than ε and ξ* measures deviations which are more than ε below the target value, as shown in Fig. 1.

The idea of SVR is to minimize an objective function which considers both the norm of the weight vector *w* and the losses measured by the slack variables (see Eq. (1)). The minimization of the norm of *w* is one of the ways to ensure the flatness of *f* [11].

The SVR algorithm involves the use of Lagrangian multipliers, which rely solely on dot products of $\phi(x)$. This can be accomplished via kernel functions, defined as $K(x_i, x_j) = \langle (x_i), (x_j) \rangle$. Thus, the method avoids computing the transformation $\phi(x)$ explicitly. The details of the solution can be found in [11].

IV. Experiments

The regression methods considered in this paper were compared using the well-known COCOMO software project dataset, reproduced in Table I .This dataset consists of two independent variables-Size and EAF (Effort Adjustment Factor) and one dependent variable-Effort. Size is in KLOC (thousands of lines of codes) and effort is given in man-months [1].In this work we are interested in estimating the effort of future projects, where the effort is given in man-months. The simulations were carried out using the Weka tool [13]. In Weka, SVR is implemented using the Sequential Minimal Optimization (SMO) algorithm [6].

TABLE I. COCOMO Dataset.

Project No.	Size	EAF	Effort
1	46	1.17	240
2	16	0.66	33
3	4	2.22	43
4	6.9	0.4	8
5	22	7.62	107
6	30	2.39	423
7	18	2.38	321
8	20	2.38	218
9	37	1.12	201
10	24	0.85	79
11	3	5.86	73
12	3.9	3.63	61
13	3.7	2.81	40
14	1.9	1.78	9
15	75	0.89	539
16	90	0.7	453
17	38	1.95	523
18	48	1.16	387
19	9.4	2.04	88
20	13	2.81	98
21	2.14	1	7.3

The following section describes the experimentation part of work, and in order to conduct the study and to establish the affectivity of the models from COCOMO dataset were used. We calculated an

Intermediate COCOMO effort by using the following equations:

$$\text{Effort} = a*(size)^{b} * EAF \qquad (2)$$

where *a* and *b* are the set of values depending on the complexity of software (for organic projects *a*=3.2, *b*=1.05, for semi-detached *a*=3.0, *b*=1.12 and for embedded *a*=2.8, *b*=1.2) and the MOPSO model effort[18]is calculated by using following equations:

$$\text{Effort} = a*(size)^{b} * EAF + C \qquad (3)$$

where *a* and *b* are cost parameters and *c* is bias factor. *a*=3.96, *b*=1.12 and *c*=5.42.The performance measures considered in our work are Mean Absolute Relative Error (MARE) and Prediction (25). The MARE is given by the following equation:

$$\text{MARE}= \frac{1}{n} \sum_{i=1}^{n} \left| fi \text{ - } yi \right| \qquad (4)$$

Pred (25) is defined as the percentage of predictions falling within 25% of the actual known value, Pred (25). *fi* is the Estimated and *yi* is the Actual value respectively, *n* is the number of data points.

We have carried out simulations considering estimating the SVR effort using both independent variables (Size and EAF). The results of our simulations are shown in Table II.

TABLE II. ESTIMATED EFFORTS OF DIFFERENT TYPES OF MODELS

Project No.	Size	EAF	Measured Effort	COCOMO Effort	MOPSO Effort	SVR Effort	COCOMO Error	MOPSO Error	SVR Error
1	46	1.17	240	208.56	342.84	239.66	31.44	102.84	0.34
2	16	0.66	33	38.82	63.74	88.51	5.82	30.74	55.51
3	4	2.22	43	30.45	46.95	42.32	12.55	3.95	0.68
4	6.9	0.4	8	9.73	19.2	43.21	1.73	11.2	35.21
5	22	7.62	107	626.11	967.39	174.9	519.11	860.39	67.9
6	30	2.39	423	271.97	432.46	174.31	151.03	9.46	248.69
7	18	2.38	321	158.41	245.41	115.85	162.59	75.59	205.15
8	20	2.38	218	176.93	275.46	126.52	41.07	57.46	91.48
9	37	1.12	201	158.85	258.53	201.34	42.15	57.53	0.34
10	24	0.85	79	76.52	123.71	135.91	2.48	44.71	56.91
11	3	5.86	73	59.43	84.85	72.27	13.57	11.85	0.73
12	3.9	3.63	61	48.49	71.43	55.16	12.51	10.43	5.84
13	3.7	2.81	40	35.52	53.59	53.51	4.48	13.59	13.51
14	1.9	1.78	9	11.17	19.88	37.39	2.17	10.88	28.39
15	75	0.89	539	336.18	449.18	391.82	202.82	89.82	147.18
16	90	0.7	453	324.32	433.52	465.13	128.68	19.48	12.13
17	38	1.95	523	284.42	459.45	219.21	238.58	63.55	303.79
18	48	1.16	387	216.23	356.28	263.17	170.77	30.72	123.83
19	9.4	2.04	88	68.64	104.78	80.45	19.36	16.78	7.55
20	13	2.81	98	132.89	202.22	105.03	34.89	104.22	7.03
21	2.14	1	7.3	7.12	14.71	38.13	0.18	7.41	30.83

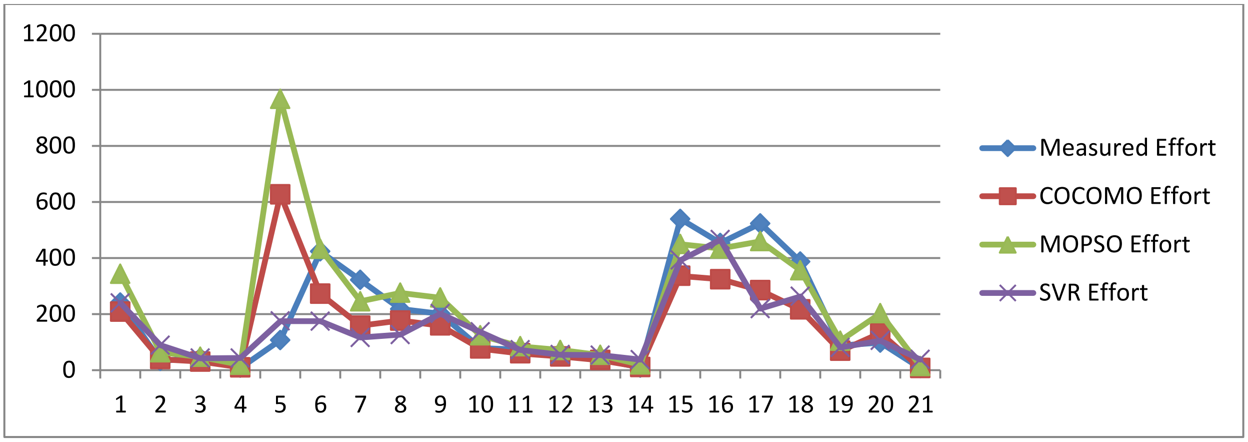


Fig.2: Measured Effort Vs Estimated Effort of various Models.

Figure 2 shows the graph of measured effort versus estimated effort of Intermediate COCOMO, MOPSO and SVR model.

From the figure 2, one can notice that the SVR estimated efforts are very close to the measured effort.

V. Results And Discussions

The results are tabulated in Table III. It was observed that the SVR gives better results in comparison with Intermediate COCOMO and MOPSO model. The MARE and Prediction accuracy is good. These results suggest that using data mining and machine learning techniques into existing software cost estimation techniques can effectively improve the accuracy of models.

Table III: Performance and Comparisons

Results	Intermediate COCOMO	MOPSO	SVR
MARE	85.62	77.74	68.72
Prediction (25%)	38.09	42.86	47.62

The following figure 3 shows the performance measures of Intermediate COCOMO, MOPSO and SVR model.

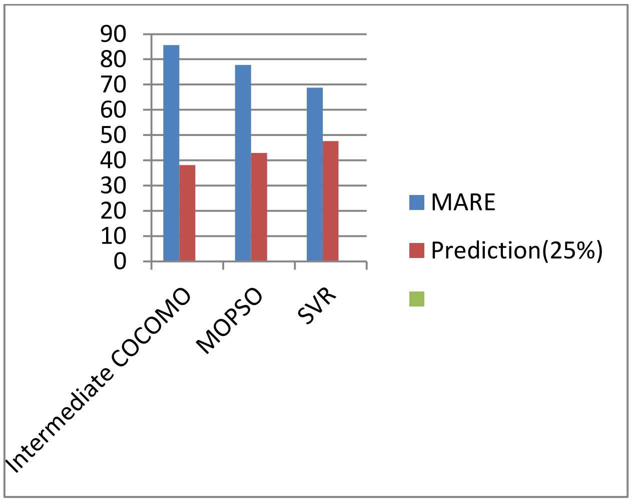


Fig.3. Performance Measure

VI. Concluding Remarks

This paper provides the use of Support Vector Regression for estimation of software project effort. We have carried out simulations using the COCOMO dataset. We have used weka tools for simulations because it consist of different-different machine learning algorithms that can be help us to classify the data easily.

The results were compared to both Intermediate COCOMO and MOPSO models. The accuracy of the model is measured in terms of its error rate. It is observed from the results that SVR gives better results. On testing the performance of the model in terms of the MARE and Prediction the results were found to be useful. The future work is the need to investigate some more data mining algorithms that can be help to improve the process of software cost estimation and easy to use.

ACKNOWLEDGMENT

The author would like to thank the anonymous referees for their helpful comments and suggestions.

REFERENCES

[1] J.W. Bailey, V.R. Basili, A meta model for software development resource expenditure, in: Proceedings of the Fifth International Conference on Software Engineering, San Diego, California, USA, 1981, pp. 107–116.

[2] B.W.Boehm, “Software Engineering Economics,” Prentice- Hall, Englewood Cliffs, NJ, USA, 1981.

[3] A.J. Albrecht and J.E. Gaffney, “Software function, source lines of code, and development effort prediction: a software science validation,” IEEE Transactions on Software Engineering, 1983, pp. 639–647.

[4] J.E. Matson, B.E Barrett and J.M. Mellichamp, “Software development cost estimation using function points,” IEEE Transactions on Software Engineering, 1994, pp. 275–287.

[5] A.R. Gray,“A simulation-based comparison of empirical modelling techniques for software metric models of development effort,” In: Proceedings of ICONIP, Sixth International Conference on Neural Information Processing, Perth, WA, Australia, 1999, pp. 526–531.

[6] G.W. Flake, S. Lawrence, Efficient SVM regression training with SMO, Mach. Learn. 46 (1–3) (2002) 271–290.

[7] A.Idri, T.M. Khosgoftaar and A. Abran, “Can neural networks be easily interpreted in software cost estimation,” World Congress on Computational Intelligence, Honolulu, Hawaii, USA, 2002, pp. 12–17.

[8] X.Huang, L.F.Capetz,J. Ren and D.Ho, “A neuro-fuzzy model for software cost estimation,” Proceedings of the third International Conference on Quality Software, 2003, pp. 126-133 .

[9] M. Lefley and M. J. Shepperd, “Using Genetic Programming to Improve Software Effort Estimation Based on General Data Sets”, LNCS, Genetic and Evolutionary Computation — GECCO 2003, ISBN: 978-3-540-40603-7, page-208.

[10] B. Kitchenham, L.M. Pickard, S. Linkman and P.W. Jones, “Modelling software bidding risks,” IEEE Transactions on Software Engineering, 2003, pp. 542–554.

[11] A.J. Smola, B. Scholkopf, A tutorial on support vector regression, Stat. Comput. 14 (3) (2004) 199–222.

[12] K.K. Aggarwal, Y. Singh, P.Chandra and M.Puri, “An expert committee model to estimate line of code,” ACM New York, NY, USA, 2005, pp. 1-4.

[13] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, second ed., Morgan Kaufmann, San Francisco, 2005.

[14] N.H. Chiu and S.J.Huang, “The adjusted analogy-based software effort estimation based on similarity distances,” System and Software, 2007, pp.628-640.

[15] K. Vinaykumar, V. Ravi, M. Carr and N. Rajkiran, “Software cost estimation using wavelet neural networks,” Journal of Systems and Software, 2008, pp. 1853-1867.

[16] Y.F. Li, M. Xie and T.N. Goh, “A study of project selection and feature weighting for analogy based software cost estimation,” Journal of Systems and Software, 2009, pp. 241–252.

[17] K. Vinay Kumar, V. Ravi and Mahil Carr, “Software Cost Estimation using Soft Computing Approaches,” Handbook on Machine Learning Applications and Trends: Algorithms, Methods and Techniques, Eds. E. Soria, J.D. Martin, R. Magdalena, M.Martinez, A.J. Serrano, IGI Global, USA, 2009.

[18] Prasad Reddy P.V.G.D, Hari CH.V.M.K and Srinivasa Rao, “Multi Objective Particle Swarm Optimization for Software Cost Estimation,” International Journal of Computer Applications, 2011, Vol.-32.

[19] Narendra Sharma and Ratnesh Litoriya, “Incorporating Data Mining Techniques on Software Cost Estimation: Validation and Improvement, ” International Journal of Emerging Technology and Advanced Engineering, 2012, vol.-2

Attribute Analysis for Bangla Words for Universal Networking Language(UNL)

Aloke Kumar Saha
Dept. of Computer Science & Engineering
University of Asia Pacific
Dhaka, Bangladesh

Muhammad Firoz Mridha
Dept. of Computer Science & Engineering
University of Asia Pacific
Dhaka, Bangladesh

Shammi Akhtar
Dept. of Computer Science & Engineering
University of Asia Pacific
Dhaka, Bangladesh

Jugal Krishna Das
Dept. of Computer Science and Engineering
Jahangirnagar University
Savar, Dhaka, Bangladesh

***Abstract*—The Universal Networking Language (UNL) is an artificial worldwide generalizes form human interactive in machine independent digital platform for defining, recapitulating, amending, storing and dissipating knowledge or information among people of different affiliations. The theoretical and practical research associated with these interdisciplinary endeavor facilities in a number of practical applications in most domains of human activities such as creating globalization trends of market or geopolitical independence among nations. In our research work we have tried to develop analysis rules for Bangla part of speech which will help to create a doorway for converting the Bangla language to UNL and vice versa and overcome the barrier between Bangla to other Languages.**

Keywords—Universal Networking Language; morphology; Bangla part of speech; morphological rules.

I. Introduction

Today the regional economics, societies, cultures and education are integrated through a globe-spanning network of communication and trade. This globalization trend evokes for a homogeneous platform so that each number of the platform can apprehend what other intimates and perpetuates the discussion in a mellifluous way. However the barriers of languages throughout the world are continuously obviating the whole world from congregating into a single domain of sharing knowledge and information. As a consequence United Nation University/Institute of advance Studies (UNU/IAS) were decided to develop an inter-language translation program. The corollary of their continuous research leads to a common form of language known as Universal Networking Language (UNL) [1].

The UNL acts as an intermediate form computer semantic language whereby any text written in a particular language is converted to a text of any other forms of language [2]. UNL, in other words is an artificial language for computer to express information and knowledge that can expressed in natural language. The rest of the paper is organized as the following. Section II outlines the UNL general structure. In Section III Bangla part of speech and in section IV provides a Rule generation for Bangla part of speech for UNL expression.

II. Structure Of UNL

UNL system composed of three parts namely Universal words, Attributes labels and relational labels. Universal word (UW) which is actually nothing but English like word and is represented by nodes in a hyper graph [1,7]. Nodes associated with a sentence are connected by a relation known as symbolic relation. Each UW has some attributes that uniquely specifies that word and is placed according to a conceptual hierarchy derives from a knowledge base. However each of the UWs is comprised of Headword along with some constraints. The headword is considered as the form of native language word known as label whereas each of constraints in a constraint list of the Universal word corresponds to a concept of that word. The attributes lists associated with the individual universal word are used to represent the subjectivity of word based on their grammatical properties [5, 6].

The knowledge base (KB) which actually holds every possible combination of semantic relations basically plays two roles. Firstly it defines semantics of UWs and then provides linguistics knowledge of concepts. The KB however not only provides linguistics knowledge in computer understandable format but also provides the semantics background of UNL expressions [8].

In addition to the above parts the UNL system has a language server which can be fragmented into two parts known as en-converter (EnCo) and de-converter (DeCo). The converter builds a framework, independent of the diversity of languages, for morphological, semantic analysis and converts the native language text into UNL expressions autonomously [14]. To perform the conversion operation the EnCo uses word dictionary, knowledge base and en-cnversion rules. In contrast the DeCo acts the reverse way the EnCo does [15]. The general formats of the word dictionary entry are defined by UNL as follows:

[HW] “UW” (ATTRIBUTE1, ATTRIBUTE2 …) <FLG, FRE, PRI>

HW← Head Word (Bangla Word)

UW← Universal Word

ATTRIBUTE← Attribute of the HW

FLG← Language Flag

FRE← Frequency of Head Word

PRI← Priority of Head Word

III. MORPHOLOGICAL ANALYSIS OF BANGLA PART OF SPEECH

Morphology is the field of linguistics that studies the structure of words. It focuses on the patterns of word formation within and across languages, and attempts to formulate rules that model the knowledge of speakers of those languages. In natural language processing (NLP) we need to identify words in texts in order to determine their syntactic and semantic properties [10, 11]. In the following section we are analyzing morphologically the different Bangla part of speech so that we can develop efficient rules for UNL expression.

A. Grammatical Construction of words

In this section, we have pointed out some essential grammatical issues about different parts of speech of Bangla that must be needed for English to Bangla MT dictionary.

B. Parts of speech

In Bangla language word may be categorized in one of five categories: noun, pronoun, adjective, verb and indeclinable [16]. Here, adverb is considered as adjective and the type indeclinable is concerned as preposition, conjunction and interjection.

Noun Morphology: Bangla Nouns have very strong and structural inflectional morphology base on case. Case of noun may be nominative ("ছেলে", boy), accusative (ছেলে-কে") , to the boy) and genitive ("ছেলে-র", of the boy) and so on. Gender and number are also important for identifying proper categories of nouns. Number may be singular ("ছেলে", boy or "ছেলেটি", the boy, "বই", book, "বইটি", the book) plural ("ছেলেরা", boys "ছেলেগুলি", the boys "বইগুলো", the books etc.). So, from the word "ছেলে" we get "ছেলের", "ছেলেকে", "ছেলেটি", "ছেলেগুলি" etc. and from the word "বই" we get "বইটি", "বইগুলি" etc. Some dictionary entries may look like.

[ছেলে]{} "boy (icl>person)" (N, HN, C, ANI)<B,0,0>

Here, "boy (icl>person)" is the UW for "ছেলে" but "র", "কে" etc. have no UWs. Therefore, they should be represented in the dictionary only with grammatical attributes as follows.

[রা] {} "" (3P, PL, SUF, N, HUMN) <B,0,0>

[টি] {} "" (N, SG, SUF,3P) <B,0,0>

[গুলি] {} "" (N, PL, SUF,3P) (<B,0,0>

[গুলো] {}""(N, SG, SUF,3P) <B,0,0>

We use 3P, SUF and N as grammatical attributes with "রা", because "রা" is used with third person say "ছেলেরা", N for noun and SUF as "রা" is a suffix. We have to put meticulous attention while defining the grammatical attributes. Because we use HUMN for human noun as "কে", "রা" are used with human being only, say ছেলেকে, তাহাকে, but not গরকে,গরুরা etc. But we can not use HUMN with "র", "টি", "¸গুলি" and "¸গুলো" because they are used with both human and non human, say পাখির, ছেলেটি, গরুগুলো, etc.

Pronoun Morphology: There are eight different types of pronouns in English language. In Bangla language, there are ten types. They are: (i) Personal (I — Avwg (pronounce as "Ami")), (ii) Reflexive (Myself — ¯^qs), (iii) Near indicating demonstrative (This — GB), (iv) Far indicating demonstrative (That — H), (v) Collective (All — mKj (pronounce as "Sakal")), (vi) Interrogative (What — wK), (vii) Indefinite (Some — wKQy), (viii) Reciprocal (Each other — ci¯úi), (ix) Relative (Who — whwb) and (x) Others indicating (Other — Ab¨). In English language, near indicating and far indicating demonstrative pronoun is collectively defined as demonstrative pronoun. Collective and others indicating pronouns are concerned as indefinite pronoun. Here we can consider the word root "তাহা"(he/she). From this we get তাহা-রা, তাহা-কে, তাহা-দের, তাহা-দের-কে, তাহা-দিগকে etc. So, we have to consider these morphemes রা, কে, দের, দিগকে for dictionary entries to form words with "তাহা" as above.

Adjective Morphology: Adjectives are of four types: proper adjective (evsjv‡`kx Kvco (pronounce as "Bangladeshi Kapor")), adjective of quality (my›`i dzj (pronounce as "Sundor Ful")), adjective of quantity (wØ¸b), pronominal adjective (‡h †Kvb †jvK Avm‡jB n‡e). In English, there are eight types of adverbial adjectives; but in Bangla, there are four categories. They are: adjective that modifies a verb (ax‡i hvI), adjective that modifies another adjective (Lye fvj †jvK (pronounce as "Khub Valo Loke")), adjective that modifies an indeclinable (Avwg †Zvgvi gZ `ª"Z Pwj), adjective that modifies a sentence (avwg©K †jv‡KivB cÖK„Z myLx). As Adjective we can consider Bangla words "সাহস", "সুন্দর" and "ভাল" meaning "bravery", "beautiful" and "good" in English respectively. From the first word we get সাহসী (সাহস+ই), সাহসের (সাহস+এর). And from the second and third words we get সুন্দরী, ভালোর, ভালোটা etc. We have to have the dictionary entries for সাহস, সুন্দর, ভালো, ই, এর, র, টা to make the meaningful words সাহসী, সাহসের, সুন্দরী, ভালোটা etc. by combining the morphemes with the root words using analysis rules.

Verb Morphology: Verbs, one of the most important word categories for dictionary design, may be classified in six categories. They are: (i) Intransitive verb (‡Q‡jiv ‡Lj‡Q (pronounce as "Chelera Khelche")), (ii) Transitive verb (‡Q‡jiv ej †Lj‡Q), (iii) Di-transitive verb (evev Avgv‡K GKwU Kjg w`‡q‡Qb), (iv) Causative verb(gv wkï‡K Pvu` †`Lvb), (v) Compound verb (a nonfinite verb + a finite verb, NUbvwU ï‡b ivL) and (vi) Complex verb (a noun/adjective/resounding indeclinable + verb, ZvRgnj `k©b

Kijvg (pronounce as “Tajmohol Dorshon Korlam”)). In verb, mood is also an important feature. There are four types: (i) Indicative Mood (Avwg eB cwo), (ii) Imperative Mood (gb w`‡q co), (iii) Subjunctive Mood (co‡j cvm Ki‡e), and (iv) Optative Mood (Zvi g½j †nvK). We can give the example of the Bangla word “চল” (means go). The corresponding UW in basic form is “go”. The dictionary entry is: [চলা] { } “go (icl>do)”, where ‘চলা&’ is the *head word* and (icl>do) is from the knowledge base. Some possible transformations of “চল" in the Bangla to UNL dictionary are given as follows [9, 10].

If we consider ‘চল’ (means go) as a root, we can represent this root in the dictionary as

[চল]{} “go (icl>do)” (V, @present) <B,0,0>

Some transformations based on the persons and tenses are.

For first person:

[চল] { } “go (icl>do)” (ROOT, BANJANT)<B, 0, 0>

[ই] { } “go (icl>do)” (ROOT, BANJANT, PRESENT INDEF)<B, 0, 0>

[ইতেছি] { } “go (icl>do)” (ROOT, BANJANT, PRESENT CONT)<B,0, 0>

For second person:

[চল] { } “go (icl>do)” (ROOT, BANJANT)<B, 0, 0>

[ইতেছ] { } “go (icl>do)” (ROOT, BANJANT, PRESENT CONT)<B, 0, 0>

[ইবে] { } “go (icl>do)” (ROOT, BANJANT, FUTURE INDEF)<B, 0, 0>

For third person:

[চল] { } “go (icl>do)” (ROOT, BANJANT)<B, 0, 0>

[ইয়াছে] { } “go (icl>do)” (ROOT, BANJANT, PRESENT PERF)<B, 0, 0>

[ইতেছে] { } “go (icl>do)” (ROOT, BANJANT, PRESENT CONT)<B, 0, 0>

For resolving the ambiguities of the words গিয়েছি, গিয়েছিলাম, গিয়েছেন, গিয়েছিলেন, যাইতে থাকবে, etc. we have to define them as full words for dictionary entries. For instance [গিয়েছিলাম] {} “go (icl>move>do)”(V, PAST, INDEF, 1P). Using the same procedure we can make dictionary entries for different transformations of other roots such as কর (do), লিখ (write),দে (give) etc.

C. Gender

In both Bangla and English Language, Gender is classified into four groups. They are: (i) Masculine (Father — evev (pronounce as “Baba”)), (ii) Feminine (Mother — gv (pronounce as “Ma”)), (iii) Common (Human — gvbyl (pronounce as “Manush”)) and (iv) Neuter (Book – eB (pronounce as “Boi”)).

D. Number

Both Bangla and English language, there are two types of number. They are: (i) Singular number, and (ii) Plural number.

IV. RULES FOR UNL TEXT GENERATION

In this section, we have presented some Bangla morphological rules for regular inflections, derivations and compounding with additional explicit rules for irregular inflection, derivation and compounding.

A. Analysis Rules

An analysis rule describes rule application conditions, a method to rewrite the attribute of node that satisfies the application condition, and construction methods of syntax tree. While applying rules, the EnConverter analyzes morphemes, syntax and semantics. Finally, it generates a syntax tree and a network.

The description format of the analysis rules is as follows [11]:

<TYPE>

[“(“<PRED>”)]…“{“[<COND1>]”:” [<ACTION1>] “:” [<RELATION1>] “:”

[<ROLE1>] “}”

“{” [<COND2>] “:” [<ACTION2>] “:” [<RELATION2>] “:”

[<ROLE2>] “}”

[“(“<SUF>”)”]…“P(“<PRIORITY>”);”

***Symbol Explanation*:**

“” represents terminal symbol,

[ ] represents zero or more times,

{} and () designates an analysis windows in the node list.

Description of Condition:

<PRE> Describes condition of nodes on the left side of the left of analysis window.

<SUF> Describes condition of nodes on the right side of the right of analysis window.

<COND1> Describes condition of the node in the Left Analysis Window (LAW).

<COND2> Describes condition of the node in the Right Analysis Window (RAW).

Description of Action:

<ACTION1> Describe the rewriting of grammatical attribute in the LAW.

<ACTION2> Describe the rewriting of grammatical attribute in the RAW.

Direction of Semantic Relation:

It describes the semantic relation between the left node (LN) and the right node (RN).

<RELATION1>Describe the semantic relation of the RAW to LAW.

<RELATION2>Describe the semantic relation of the LAW to RAW.

<PRIORITY> Describes priority of the rules. Code 0-255 is used to specify the priority.

B. *Types of the Analysis Rules*

This part explains the action and functions of the rule types that can lie described with <TYPE> in analysis rules.

Left Composition " + | +:+ | +: c | +:*"

The RN is combined to LN to make one composition node. The syntax tree and the attribute having left node are inherited. When the RN attributes is inherited, "@" is put in the action column of the LN, the original two nodes are deleted from the node list. The composition node is inserted into the node list. After applying the rules, the composition node takes a position in the RAW.

Right Composition " - | -:+ | -:c | -:* "

The LN is combined to RN to make one composition node. The composition node is inserted into the node list. After applying the rules, the composition node takes a position in the LAW.

Left Modification "<"

When the RN modifies LN, the RN is deleted from node list and the LN remains only. The node, which the <RELATION> is described, is the to-node and the other node is from-node.

Right Modification ">"

When the LN modifies RN, the LN is deleted from node list and the RN remains only.

Left Shift "L"

Shift the analysis window to the left.

Right Shift "R"

Shift the analysis window to the right.

Attribute Changing Rule ":"

This rule adds or deletes attributes from a particular node.

C. *Morphological rule Generation for Bangla Parts of Speech*

Bangla is a semantic language, and its basic characteristic is the rich morphology in which most of its words are derived from roots. Inflections and derivations are generated by changing vowels and insertion of consonants. Bangla sentences are characterized by a strong tendency for agreement between its constituents: between verb and noun, noun and objective, in matters of numbers, gender, definitiveness, case, person, etc.

These properties are expressed by a comprehensive system of affixation. To satisfy these grammatical properties, the generation rules are expected to be complex for handling the processing of generating grammatically correct Bangla sentences from UNL expression and structure A database system has been developed for the classification and features adding for each entry in the dictionary [3, 4]. The selected Bangla word is then classified to Noun or Verb or Particle. The relation mapping is implemented in the en-conversion rule.

Bangla parts of speech conversion rules are mainly for noun ↔ adjective. Some conversion rules are also done for indeclinable ↔ noun and indeclinable ↔ adjective and there are some exceptions also.

From the analysis of Bangla Part of speech, gender and number, one can readily find that they agree right composition rule.

a) Right Composition Rule: (For Bangla noun ↔ adjective)

- Rule 1: C {NOUN:::} { CASEMARKER : +ADJ,-CASEMARKER::}
  For noun ↔ adjective, add ' x ' with the last character (AvMgb–AvMgbx).

- Rule 2: C {NOUN:::} { CASEMARKER : +ADJ,-CASEMARKER::}
  For noun ↔ adjective, add 'w ' with the last character and after that, add ' Z '(Avb›` – Avbw›`Z, Av‡gv` –Av‡gvw`Z).

- Rule 3: C {NOUN:::} { CASEMARKER : +ADJ,-CASEMARKER::}
  For noun ↔ adjective, add '‡U' after the last character (SMov –SMov‡U).

- Rule 4: C {NOUN:::} { CASEMARKER : +ADJ,-CASEMARKER::}
  For noun ↔ adjective, add ' x ' with the last character and after that, add ' q '(bvUK – bvUKxq)

- Rule 5: C {NOUN:::} { CASEMARKER : +ADJ,-CASEMARKER::}
  For noun ↔ adjective, if the last character is 'b / Y / q ', then drop the last character and add 'w ' with the previous character and after that, add ' Z '(mvab – mvwaZ, AvniY – AvnwiZ, cwiPq – cwiwPZ).

- Rule 6: C {NOUN:::} { CASEMARKER : +ADJ,-CASEMARKER::}
  For noun ↔ adjective, add '¨' after the last character (‡fvM –‡fvM¨).

- Rule 7: C {NOUN:::} { CASEMARKER : +ADJ,-CASEMARKER::}
  For noun ↔ adjective, add ' D ' with the last of the word (Xvj + D = Xvjy).

- Rule 8: C {NOUN:::} { CASEMARKER : +ADJ,-CASEMARKER::}

For noun ↔ adjective, add ‘DK ’ at last (‡cU + DK = ‡cUzK).

- Rule 9: C {NOUN:::} { CASEMARKER : +ADJ,-CASEMARKER::}
  For noun ↔ adjective, add ’Av ’ at last (evN + Av = evNv).

- Rule 10: C {NOUN:::} { CASEMARKER : +ADJ,-CASEMARKER::}
  For noun ↔ adjective, add ‘ Bqv / D‡i ’ at last (kni + Bqv/D‡i = knwiqv/kû‡i).

- Rule 11: C {NOUN:::} { CASEMARKER : +ADJ,-CASEMARKER::}
  For noun ↔ adjective, add ‘†P ’ at last (jvj + ‡P =jvj‡P).

- Rule 12: C {NOUN:::} { CASEMARKER : +ADJ,-CASEMARKER::}
  For noun ↔ adjective, add ‘‡Zv ’ at last (dzcv + †Zv =dzcv‡Zv).

- Rule 13: C { ADJ :::} { CASEMARKER : + NOUN ,- CASEMARKER::}
  For adjective ↔ noun, add ‘AvB ‘ at last (wgVv + AvB =wgVvB).

- Rule 14: C { ADJ :::} { CASEMARKER : + NOUN ,- CASEMARKER::}
  For adjective ↔ noun, add ‘Bgv ’ at last (bxj + Bgv =bxwjgv).
- Rule 15: C { ADJ :::} { CASEMARKER : + NOUN ,- CASEMARKER::}
  For adjective ↔ noun, add ‘cbv ‘ at last (`~iš— + cbv = `~iš—cbv).

- Rule 16: C {ADJ:::} { CASEMARKER : +NOUN,-CASEMARKER::}
  For adjective ↔ noun, add ‘ Avwg ’ at last (`~ó + Avwg =`~óvwg).

- Rule 17: C {ADJ:::} { CASEMARKER : +NOUN,-CASEMARKER::}
  For adjective ↔ noun, add ‘wMwi’ at last (evey + wMwi =eveywMwi).

- Rule 18: C {ADJ:::} { CASEMARKER : +NOUN,-CASEMARKER::}
  For adjective ↔ noun, add ‘ Zv ’ at last (GK + Zv = GKZv).

- Rule19: C{INDECLINABLE:::} {CASEMARKER: +NOUN,- CASEMARKER::}
  For indeclinable ↔ noun, drop ‘ v ’ and add ‘ ¨ ’ (Z_v –Z_¨).

b) Rules for Gender

In gender, a masculine or a feminine word is derived from another feminine or another masculine word. Here we have derived some morphological rules for the conversion of Bangla Gender:

- Rule 1: C { MALE :::} {CASEMARKER : +FEMALE,-CASEMARKER,+ CASEMARKER::}
  Dropping the last ‘ v ’ and adding the vowel ‘ x ’, some masculine words are converted to feminine (PvPv – PvPx) .
- Rule 2: C { MALE :::} {CASEMARKER : +FEMALE,-CASEMARKER,+CASEMARKER::}
  Adding ‘bx’, some masculine words are converted to feminine (‡avcv – ‡avcvbx).
- Rule 3: C { MALE :::} {CASEMARKER : +FEMALE,- CASEMARKER,+ CASEMARKER::}
  If there is ‘ x ’ in the masculine word, then the ‘ x ’ is converted to ‘ w ’ and at the end, ‘ bx ’ is added (wfLvix — wfLvwibx).

- Rule 4: C { MALE :::} {CASEMARKER : +FEMALE,- CASEMARKER,+ CASEMARKER::}
  Adding ‘Avbx’ , some masculine words are converted to feminine (wng – wngvbx).

c) Rules for Number

Here are few rules for the translation of Bangla Number:

- Rule 1: C {NOUN, SING:::} {NUMBERSIGN : +PLU,- NUMBERSIGN ::}
  Adding “ wU, Uv, Lvbv, Lvwb” with the main word represents singular word (Mi“ – Mi“wU, evQyi – evQyiUv, LvZv – LvZvLvbv, eB–eBLvwb).

- Rule 2: C {NOUN,SING,HIGH:::} {NUMBERSIGN : +PLU,- NUMBERSIGN ::}
  Adding “iv, Giv, MY, e„›`, gÛjx, eM©” with the main word represents the high class living things in plural number (QvÎ –QvÎiv, gv–gv‡qiv, Rb– RbMY, wk¶K – wk¶KgÛjx .

- Rule 3: C {NOUN,SING,LIVING:::} {NUMBERSIGN : +PLU,- NUMBERSIGN ::}
  Adding “ Kzj, mKj, me, mg~n ” with the main word represents the low class living things and non-living things in plural number (Kwe–KweKzj, fvB–fvBme).

- Rule 4: C {NOUN, SING, NON-LIVING :::} {NUMBERSIGN: +PLU,- NUMBERSIGN ::}
  Adding “ ¸jv, ¸wj, ¸‡jv, ¸”Q, `vg, ivwk, Avewj, wbKi, cyÄ, gvjv, ivwR, wbPq ” with the main word represents the non-living things and dumb things in

plural number (Avg–Avg¸jv,Kzmyg–Kzmyg`vg, evwj–evwjivwk, ZviKv –ZviKvivwR).

V. Conclusion

In this paper we have presented morphological rules for Bangla part of speech, number and gender. To do so we did morphological analysis of Bangla part of speech. We hope that these rules would be useful for conversion of Bangla sentence to UNL expressions and vice-versa.

Even though the limited numbers of rules are considered in this paper, it theoretically shows that the designed model works perfectly for Bangla words. All the Bangla words and rules will be considered in future.

References

[1] H. Uchida, M. Zhu, “The Universal Networking Language (UNL) Specification Version 3.0”, Technical Report, United Nations University, Tokyo, 1998

[2] Uchida H.,Zhu,M. and Della Senta, “A Gift for a Millennium. The United Nations University,Tokyo,Japa,2000”

[3] Muhammad Firoz Mridha, Manoj Banik, Md. Nawab Yousuf Ali, Mohammad Nurul Huda, Chowdhury Mofizur Rahman, Jugal Krishna Das, “Formation of Bangla Word Dictionary Compatible with UNL Structure”, SKIMA 2010, August 25-27, Paro, Bhutan.

[4] Muhammad Firoz Mridha, Md. Zakir Hossain, Manoj Banik, Mohammad Nurul Huda, ChowdhuryMofizur Rahman, Jugal Krishna Das, “Development of Grammatical Attributes for Bangla Root and Primary Suffix for Universal Networking Language”, SKIMA 2010, August 25-27, Paro, Bhutan.

[5] S. Abdel-Rahim, A.A. Libdeh, F. Sawalha, M. K. Odeh, “Universal Networking Language(UNL) a Means to Bridge the Digital Divide”, Computer Technology Training and Indistrial Studies Center, Royal Scientific Sciety, March 2002.

[6] M. M. Asaduzzaman, M. M. Ali, “Morphological Analysis of Bangla Words for Automatic Machine Translation”, International Conference on Computer, and Information Technology (ICCIT), Dhaka, 2003, pp.271-276

[7] Serrasset Gilles, Boitel Christian, (1999) UNL-French Deconversion as Transfer & Generation from an Interlingua with Possible Quality Enhancement through Offline Human Interaction. *Machine Translation Summit-VII*, Singapore.

[8] M. E. H. Choudhury, M. N.Y. Ali, M.Z.H. Sarkar, R. Ahsan, “Bridging Bangla to Universal Networking Language- A Human Language Neutral Meta- Language”, International Conference on Computer and Information Technology (ICCIT), Dhaka, 2005,pp.104- 109

[9] M.E.H. Choudhury, M.N.Y. Ali, “Framework for synthesis of Universal Networking Language”, East West University Journal, Vol. 1, No. 2, 2008, pp. 28-43

[10] M.N.Y. Ali, J.K. Das, S.M. Abdullah Al Mamun, M. E.H. Choudhury, ”Specific Features of a Converter of Web Documents from Bengali to Universal Networking Language”, International Conference on Computer and Communication Engineering 2008(ICCCE’08), Kuala Lumpur, Malaysia.pp. 726-731

[11] S. Dashgupta, N. Khan, D.S.H. Pavel, A.I. Sarkar, M. Khan, ”Morphological Analysis of Inflecting Compound words in Bangla”, International Conference on Computer, and Communication Engineering (ICCIT), Dhaka, 2005, pp. 110-117

[12] M.N.Y. Ali, J.K. Das, S.M. Abdullah Al Mamun, A. M. Nurannabi,” Morpholoical Analysis of Bangla worfs for Universal Networking Language”,icdim,08.

[13] Bangla Academy (2004), Bengali-English Dictionary, Dhaka.

[14] Enconverter Specifications, version 3.3, UNL Center/ UNDL Foundation, Tokyo, Japan 2002.

[15] Deconverter Specifications, version 2.7, UNL Center/ UNDL Foundation, Tokyo, Japan, 2002

[16] D. S. Rameswar, “Shadharan Vasha Biggan and Bangla Vasha”, Pustok Biponi Prokashoni, November 1996, pp.35

[17] D. S. Rameswar, “Shadharan Vasha Biggan and Bangla Vasha”, Pustok Biponi Prokashoni, November 1996, pp.35

Shadow Suppression using RGB and HSV Color Space in Moving Object Detection

Shailaja Surkutlawar
EXTC
Vivekananda education society of information technology
Mumbai, India

Prof. Ramesh K Kulkarni
EXTC
Vivekananda education society of information technology
Mumbai, India

***Abstract*— Video-surveillance and traffic analysis systems can be heavily improved using vision-based techniques to extract, manage and track objects in the scene. However, problems arise due to shadows. In particular, moving shadows can affect the correct localization, measurements and detection of moving objects. This work aims to present a technique for shadow detection and suppression used in a system for moving visual object detection and tracking. The major novelty of the shadow detection technique is the analysis carried out in the HSV color space to improve the accuracy in detecting shadows. This paper exploits comparison of shadow suppression using RGB and HSV color space in moving object detection and results in this paper are more encouraging using HSV colour space over RGB colour space.**

Keywords— Shadow detection; HSV color space; RGB color space.

I. Introduction

Surveillance systems have wide demand in public areas, such as airports, subways, entrance to buildings. In this context, reliable detection of moving objects is most critical requirement for the surveillance systems. To detect a moving object, a surveillance system usually utilizes background subtraction. The key of background subtraction is the background model. In the moving object detection process, one of the main challenges is to differentiate moving objects from their cast shadows.

Moving cast shadows are usually misclassified as part of the moving object making the following stages, such as object classification or tracking, to perform inaccurate. The Gaussian mixture model (GMM) [1] represented the statistics of one pixel over time can cope with multi-modal background distributions. However, a common problem for this approach is to find the right balance between the speed at which the model adapts to changing background, and the stability.

The shadow points and the object points share two important visual features: motion model and detectability. Since the most common techniques for foreground object detection in dynamic scene are inter-frame difference or background suppression, all the moving points of both objects and shadows are detected at the same time.

Moreover, shadow points are usually adjacent to object points and with the more commonly used segmentation techniques shadows and objects are merged in a single blob. These aspects cause two important drawbacks: The former is that the object shape is falsified by shadows and all the measured geometrical properties are affected by an error (that varies during the day and when the luminance changes). This affects both the classification and the assessment of moving object position (normally given by the shape centroid), as, for instance, in traffic control systems that must evaluate the trajectories of vehicles and people on a road. The second problem is that the shadows of two or more objects can create a false adjacency between one another, thus detecting them as merged in a single affects many higher level surveillance tasks such as counting and classifying individual objects in the scene. In order to avoid the drawbacks due to shadows, a new technique of shadow suppression using HSV colour space is proposed.

The paper is organized as follows: Section II deals with the background subtraction using Gaussian Mixture Model to classify the pixels as background or foreground by thresholding the difference between the background image and the current image, Section III deals with Post processing techniques for suppressing shadow using HSV and RGB colour space, Section IV discusses the experimental results of shadow suppression techniques. Finally, the conclusion is given in Section V.

II. Background Substraction

In the model of Mixture of Gauss **[1] [4] [5],** the background is not a single frame without any moving objects. Gaussian Mixture Model (GMM) is thought to be one of the best background modeling methods and works well when gradual changes appear in the scene **[2] [3]** . The GMM method models the intensity of each pixel with a mixture of $\boldsymbol{k}$ Gaussian distributions. The probability that a certain pixel has a value $\boldsymbol{X_t}$ at time can be written as

$$\mathbf{P}(\boldsymbol{X_t}) = \sum_{i=1}^{k} \boldsymbol{\omega_{i,t}} \,.\, \boldsymbol{\eta}\,(\boldsymbol{X_t}, \boldsymbol{\mu_{i,t}}, \boldsymbol{\Sigma_{i,t}}) \qquad \mathbf{(1)}$$

Where $\boldsymbol{k}$ is the number of distributions (currently, 3 to 5 is used), $\boldsymbol{\omega_{i,t}}$ is the weight of the $\boldsymbol{k}^{\mathbf{th}}$ Gaussian in the mixture at time $\boldsymbol{t}$ and $\boldsymbol{\eta}\,(\boldsymbol{X_t},\ \boldsymbol{\mu_{i,t}}\,,\boldsymbol{\Sigma_{i,t}})$ the Gaussian probability density function.

$$\boldsymbol{\eta}\,(\,X_t\,,\,\mu_{i,t}\,,\Sigma_{i,t})=\frac{1}{(2\pi)^{3/2}\,|\Sigma_{i,t}|^{1/2}}\,e^{\left\{\frac{-1(X_t-\mu_{i,t})^T\,\Sigma_{i,t}^{-1}(X_t-\mu_{i,t})}{2}\right\}} \qquad \textbf{(2)}$$

Where, $\boldsymbol{\mu_{i,t}}$ is the mean value and $\boldsymbol{\Sigma_{i,t}}$ is the covariance of the $\boldsymbol{k}^{\text{th}}$ Gaussian at time $\boldsymbol{t}$*.* For computational reasons, the covariance matrix is assumed to be of the form

$$\textstyle\sum_{k,t} = \sigma^2 . I \qquad \textbf{(3)}$$

Where $\boldsymbol{\sigma}$ is the standard deviation. This assumes that the red, green, and blue pixel values are independent and have the same variance, allowing us to avoid a costly matrix inversion at the expense of some accuracy.

Thus, the distribution of recently observed values of each pixel in the scene is characterized by a mixture of Gaussians. A new pixel value will, in general, be represented by one of the major components of the mixture model and used to update the model.

However, it fails when there are sharp changes, such as sudden illumination changes or sudden partial changes in the background. To tackle this problem, some improvement has been made in recent researches. In [6], every frame is processed on pixel level, region level and frame level with color and gradient information to overcome the problem caused by sudden illumination changes based on GMM. In [7], a hierarchical GMM using state models without temporal correlation on different scales is proposed to handle sharp changes. Zivkovic presented an improved GMM algorithm automatically fully adapting to the scene, by choosing the number of components for each pixel in an online procedure [8] [9], which leads to big improvement in reduced processing time and slight improvement in segmentation result.

If the pixel process could be considered a stationary process, a standard method for maximizing the likelihood of the observed data is expectation maximization. Unfortunately, each pixel process varies over time as the state of the world changes, therefore an approximate method which essentially treats each new observation as a sample set of size 1 and uses standard learning rules to integrate the new data.

If lighting changes occurred in a static scene, it would be necessary for the Gaussian to track those changes. If a static object was added to the scene and was not incorporated in to the background until it had been there longer than the previous object, the corresponding pixels could be considered foreground for arbitrarily long periods. This would lead to accumulated errors in the foreground estimation, resulting in poor tracking behavior. These factors suggest that more recent observations may be more important in determining the Gaussian parameter estimates. Since there is a mixture model for every pixel in the image, implementing an exact Expectation maximization algorithm on a window of recent data would be costly.

Instead, we implement an on-line K-means approximation. Every new pixel value $\boldsymbol{X_t}$**,** is checked against the existing $\boldsymbol{k}$ Gaussian distributions, until a match is found. A match is defined as a pixel value within 2.5 standard deviations of a distribution1. GMM algorithm can be summarized as:

- Initialize each pixel of the scene with $\boldsymbol{k}$ Gaussian distributions
- Every new pixel value $\boldsymbol{X_t}$**,** is checked against the existing Gaussian distributions until a match is found.
- A match is defined as a pixel value within **2.5** standard deviations of a distribution.
- If none of $\boldsymbol{k}$-distributions match current pixel value, least probable distribution is go out.
- A new distribution with current value as mean value, an initially high variance, and low prior weight, is entered.
- Prior weights of $\boldsymbol{k}$ distributions at time adjusted as

$$\boldsymbol{\omega_{k,t} = (1-\alpha)\omega_{k,t-1} + \alpha(M_{k,t})} \qquad (4)$$

- $\boldsymbol{\mu_t}$ and $\boldsymbol{\sigma}$ for unmatched distributions remain the same.
- Parameters of distribution matching new observation are updated as :

$$\boldsymbol{\mu_t = (1-\rho)\mu_{t-1} + \rho(X_t)} \qquad (5)$$

$$\boldsymbol{\sigma_t^2 = (1-\rho)\sigma_{t-1}^2 + \rho(X_t-\mu_t)^T(X_t-\mu_t)} \qquad (6)$$

$$\boldsymbol{\rho = \alpha\eta\ (X_t \mid \mu_{t-1}.\sigma_{t-1})} \qquad (7)$$

- Gaussians are ordered by the value of $\boldsymbol{\omega/\sigma}$
- 1st B distributions are chosen as background model, where

$$\boldsymbol{B = arg_b^{min}(\textstyle\sum_{k=1}^{b}\omega_k > T)} \qquad (8)$$

III. Shadow Suppression Technique

Shadows are due to the occlusion of light source by an object in the scene. In particular, that part of the object not illuminated is called self-shadow, while the area projected on the scene by the object is called cast shadow [10]. This last one is more properly called moving cast shadow if the object is moving.

A. Normalized RGB color space

The Normalized RGB space aims to separate the chromatic components from the brightness component. The red, green and blue channel can be transformed to their normalized counterpart by using the formulae

$$\boldsymbol{l = R + G + B,\ r = R/l,\quad g = G/l,\quad b = B/l} \qquad (9)$$

When $\boldsymbol{l \neq 0}$ and $\boldsymbol{r = g = b = 0}$ otherwise**.**

One of these normalized channels is redundant, since by definition ***r*, *g*,** and ***b*** sum up to 1. Therefore, the Normalized RGB space is sufficiently represented by two chromatic components ***r*** and ***g*** and a brightness component l. Normalized RGB suffers from a problem inherent to the normalization namely noise sensor or compression noise at low intensities results in unstable chromatic components.

Under the consideration of saving computational cost, RGB space based method proposed by Horprasert in **[4]** is adopted**.** The basic idea in **[4]** is that shadow has similar chromaticity but lower brightness. For a given observed pixel value I_i, a brightness distortion, α_i, and a color distortion CD_i, is calculated by,

$$\alpha_i = arg_{\alpha_i} min(I_i - \alpha_i E_i)^2 \quad (10)$$

$$CD_i = \|I_i - \alpha_i E_i\| \quad (11)$$

Where E is the expected chromaticity line , α_i equals 1 if the brightness of the given pixel in the current frame is the same as in the background image. α_i, is less than 1 if it is darker and greater than 1 if it becomes brighter than the expected brightness. Then, the criteria for shadow pixels simply becomes,

$$\begin{cases} \tau_a < \alpha_i < 1 \\ CD_i < \tau_{CD} \end{cases} \quad (12)$$

In **[11]**, τ_a and τ_{CD} are predefined thresholds $\tau_a = 0.7$ and $\tau_{CD}=5$, in our experiments.

A. HSV color space

In literature, many works have been published on shadow detection topic. Jiang and Ward **[10]** extract both self-shadows and cast shadows from a static image. They use a three level processes approach:

1. The low level process extracts dark regions by thresholding input image.
2. The middle level process detects features in dark regions, such as the vertexes and the gradient of the outline of the dark regions and uses them to further classify the region as penumbra (part of the shadow where the direct light is only partially blocked by the object), self-shadow or cast shadow.
3. The high level process integrates these features and confirms the consistency along the light directions estimated from the lower levels.

It addresses the problem of segmentation of moving objects, hence an approach for detecting moving cast shadows on the background, without computing static shadows is defined .In **[12]**, the authors detail the shadow handling system using signal processing theory. Thus, the appearance of a point belonging to a cast shadow can be described as:

$$S_k(x, y) = E_k(x, y)\rho_k(x, y) \quad (13)$$

Where S_k is the image luminance of the point of coordinate ***(x, y)*** at time instant ***t***. $E_k(x, y)$ is the irradiance and it is computed as follows:

$$E_X\,(x, y) = \begin{cases} C_A + C_P \; cos \angle(N(\,x\,, y), L) & \textbf{\textit{illuminate}} \\ C_A & \textbf{\textit{shadowed}} \end{cases} \quad (14)$$

Where C_A and C_P are the intensity of the ambient light and of the light source, respectively, ***L*** the direction of the light source and ***N (x, y)* i**s object surface normal. $\rho_k(x, y)$ is the reflectance of the object surface. In [**12],** some hypotheses on the environment are outlined:

I. strong light source
II. static background (and camera)
III. planar background

Most of the papers take implicitly into account these hypotheses. In fact, typically the first step computed for shadow detection is the difference between the current frame and a reference image, as in **[12],** or a reference frame, typically named background model **[13][14][15][16].** Difference $D_k(x,\ y)$ can be written as:

$$D_k(x, y) = S_{k+1}(x, y) - S_k(x, y) \quad (15)$$

Let us consider that a previously illuminated point is covered by a cast shadow at frame $k + 1$. According to the hypothesis **2** in **[12]** of a static background, reflectance ρ_k(x, y) of the background does not change with time, thus we can assume that

$$\rho_{k+1}(x, y) = \rho_k(x, y) = \rho(x, y) \quad (16)$$

$$D_k\,(x\,,\ y) = \rho(x\,, y)\ C_P \; cos \angle\,(N(x\,, y)\,, L) \quad (17)$$

Thus, if hypothesis 1 in [12] holds, Cp in eq.17 is high. Summarizing, if hypotheses 1 and 2 in [12] hold, difference in eq. 6 is high in presence of cast shadows covering a static background. This implies that shadow points can be obtained by thresholding the frame difference image eq. 17 detects not only shadows, but also foreground points. In [13] Kilger uses a background suppression technique to find the moving objects and moving cast shadows in the scene. Then, for each object, it exploits the information on date, time and heading of the road computed by its system to choose whether to look for vertical or horizontal edges to separate shadows from objects.

In **[17]**, a statistical posterior estimation of the pixel probabilities of membership to the class of background, foreground or shadow points, authors use three sources of information: local, based on the assumption that the appearance of a shadowed pixel can be approximated using a linear transformation of the underlying pixel appearance, according with the fact that the difference of eq. 17 should be positive; spatial, which iterates the local computation by re-computing the a-priori probabilities using the a-posteriori probabilities of the neighborhood; temporal, which predicts the position of shadows and objects from previous frames, therefore adapting the a-priori probabilities. The approach in **[12]** exploits the local appearance change due to shadow by computing the ratio $R_k(x, y)$ between the appearance of the pixel in the actual frame and the appearance in a reference frame:

$$\boldsymbol{R_k(x,y)} = \frac{\boldsymbol{S_{k+1}(x,y)}}{\boldsymbol{S_k(x,y)}} \tag{18}$$

That can be rewritten as ratio between irradiance and reflectance by using eq. 13 and eq. 16 as

$$\boldsymbol{R_k(x,y)} = \frac{\boldsymbol{E_{k+1}(x,y)}}{\boldsymbol{E(x,y)}} \tag{19}$$

If static background point is covered by a shadow, we have:

$$\boldsymbol{R_k(x,y)} = \frac{\boldsymbol{C_A}}{\boldsymbol{C_A + C_P \cos \angle (\mathrm{N}\ (x,y), \mathrm{L})}} \tag{20}$$

This ratio is less than 1. In fact, the angle between ***N(x, y)*** and $\boldsymbol{L}$ is in range between $\boldsymbol{{}^{-\pi}/_{2}}$ to $\boldsymbol{{}^{\pi}/_{2}}$ therefore the Cos function is always positive. Moreover, due to hypothesis 3, we can assume ***N(x, y)*** as spatially constant in a neighborhood of the point, as background is supposed planar in neighborhood.

In **[12],** authors exploit the spatial constancy of ***N*** to detect shadows by computing the variance in a neighborhood of the pixel of the ratio ***R_k (x, y):*** a low variance means that assumption 3 holds, then they mark that pixel as “possible shadow”, eq. 20 can be seen as the ratio between the luminance after and before shadow appears. In a similar way, Davis et al. **[11] [14]** define a local assumption on the ratio between shadow and shadowed point luminance. This is based on the hypothesis that shadows darken the covered point, as eq. 20 and the considerations above confirm.

IV. Experimental Results

Original frames for the experiments showed in Fig. a, Fig. b, Fig. c. On this original frames Background subtraction using GMM is applied, as a result background pixels , foreground pixels and some shadow pixels (falsely segmented as foreground pixels) shown as black and white respectively in Fig. d, Fig. e, Fig. f . Post processing techniques for shadow suppression using HSV and RGB color space applied on Fig. d, Fig. e, Fig. f, results shown in Fig. g, Fig. h, Fig. i using RGB color space and Fig. j, Fig. k, Fig. l using HSV color space. Results show that shadow suppression is better using HSV as compared to RGB color space.

V. Conclusion

Moving objects detection and segmentation is a fundamental step in many applications based on vision. Mixture of Gaussians is the frequently used method to subtracting moving objects from background. But its results are not good enough in some cases. In this paper, a post-processing method is proposed to solve this problem. The results with more complete boundaries provided by the color clustering is used to verify the outputs of mixture of Gaussians, and thus two possible false segmentations can be corrected effectively. Moving shadow suppression using RGB and HSV colour spaces and small region median filter are also adopted. This paper compare shadow suppression results using RGB and HSV colour space and found that results of HSV are good over RGB colour space.

References

[1] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Ft. Collins, 1999: pp.246-252.

[2] C. Stauffer, and W. E. L Grimson, “Learning patterns of activity using real-time tracking,” IEEE Trans. on PAMI, Vol. 22, No. 8,2000 pp. 747-757.

[3] P. Kumar, and A. Mittal,“Study of robust and intelligent surveillance in visible and multimodal framework,” informatica, Vol. 31, 2007 pp. 447-461.

[4] Horprasert T, Harwood D, Davis L S. "A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection,” in Proceedings of IEEE ICCV' 99 Frame-Rate Workshop, 1999, pp.1-19.

[5] KaewTraKulPong P., Bowden R, “An Improved Adaptive Background Mixture Model for Real time Tracking with Shadow Detection,” in proceedings of 2nd European Workshop on Advanced Video Based Surveillance Systems, Sept 2001, Pages:1-5.

[6] O.Javed, K.Shafique, and M .Shah., “A hierarchical approach to robust background subtraction using color and gradient information,” In Workshop on Motion and Video Computing, Dec. 2002, pp. 22–27.

[7] Y.D.Sun, and B. Z. Yuan, “Hierarchical GMM to handle sharp changes in moving object detection,” in Electronics Letters, Vol. 40, No. 13, 2004, pp. 801-802.

[8] Z.Zivkovic, “Improved adaptive gaussian mixture model for background subtraction,” in proceedings ICPR, 2004.

[9] Z. Zivkovic and F.vander Heijden, “ Efficient adaptive density estimation per image pixel for the task of background subtraction,” Pattern Recognition Letters, Vol. 27,No. 7, 2006, pp. 773-780.

[10] C. Jiang and M.O. Ward, “Shadow identification,” in Proceedings of IEEE Int’l Conference on Computer Vision and Pattern Recognition, , 1992 ,pp. 606–612.

[11] I. Haritaoglu, D. Harwood, and L.S. Davis, “W4: Real-time surveillance of people and their activities,”in proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 8, Aug. 2000, pp. 809–830.

[12] J. Stauder, R. Mech, and J. Ostermann, “Detection of moving cast shadows for object segmentation,”in proceedings of IEEE Transactions on Multimedia, vol. 1, no. 1, Mar. 1999, pp. 65–76.

[13] M. Kilger, “A shadow handler in a video-based real-time traffic monitoring system,” in Proceedings of IEEE Workshop on Applications of Computer Vision, 1992, pp. 11–18.

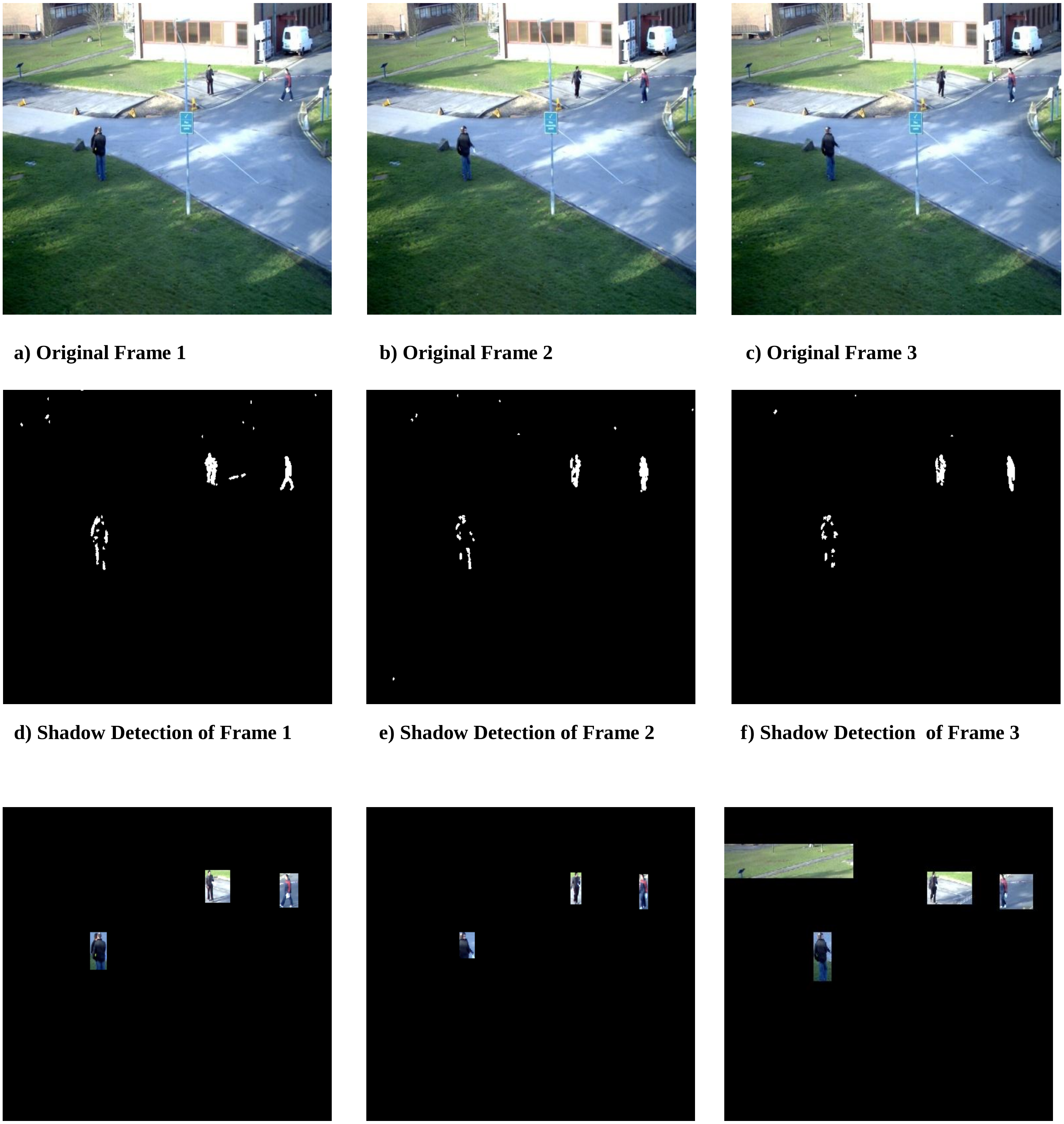
[14] A. Elgammal, D. Harwood, and L.S. Davis, “Non-parametric model for background subtraction,” in Proceedings of IEEE ICCV’99 FRAME-RATE Workshop, 1999.

[15] T. Horprasert, D. Harwood, and L.S. Davis, “A Statistical approach for real-time robust background subtraction and shadow detection,” in Proceedings of IEEE ICCV’99 FRAME-RATE Workshop, 1999.

[16] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, “Statistical and knowledge-based moving object detection in traffic scene,” in Proceedings of IEEE Int’l Conference on Intelligent Transportation Systems, Oct. 2000, pp. 27–32.

[17] I. Mikic, P. Cosman, G. Kogut, and M.M. Trivedi, “Moving shadow and object detection in traffic scenes,” in Proceedings of Int’l Conference on Pattern Recognition, Sept. 2000.

[18] N. Herodotou, K.N. Plataniotis, and A.N. Venetsanopoulos, “A color segmentation scheme for object-based video coding,” in Proceedings of the IEEE Symposium on Advances in Digital Filtering and Signal Processing, 1998, pp. 25–29.



a) Original Frame 1

b) Original Frame 2

c) Original Frame 3

d) Shadow Detection of Frame 1

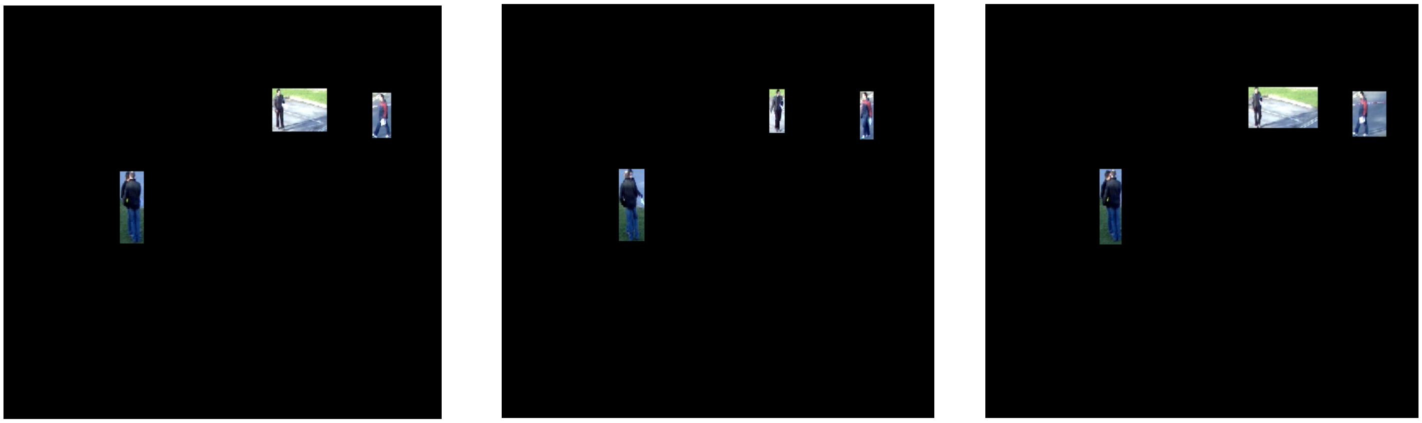
e) Shadow Detection of Frame 2

f) Shadow Detection of Frame 3

j) HSV Result Of Frame 1

k) HSV Result Of Frame 2

l) HSV Result Of Frame 3



g) RGB Result Of Frame 1 **h) RGB Result Of Frame 2** **i) RGB Result Of Frame 3**

A Proposed Integrated Approach for BI and GIS in Health Sector to Support Decision Makers (BIGIS-DSS)

Torky Sultan
Information System Dept.
Faculty of Computer and Information
Systems, Helwan University,
Cairo, Egypt

Mona Nasr
Information System Dept.
Faculty of Computer and Information
Systems, Helwan University,
Cairo, Egypt

Ayman Khedr
Information System Dept.
Faculty of Computer and Information
Systems, Helwan University,
Cairo, Egypt

Randa Abdou
Information System Dept.
Faculty of Computer and Information
Systems, Helwan University,
Cairo, Egypt

***Abstract*— This paper explores the possibilities of adopting Business Intelligence (BI), and Geographic Information System (GIS) to build a spatial intelligence and predictive analytical approach. The proposed approach will help in solving spatial problem which faces decision makers at health sector. The proposed spatial analytical approach will cover three main health planning issues. These issues are tackling health inequalities through geospatial monitor for inequalities in distribution of health units and its services, support decision-making with predictive analytics for common health indicators, and geoprocessing for input layers through dynamic health map and motion charts to support decision making.**

***Keywords*— Business Intelligence (BI); Geographic Information System (GIS); Decision Support System (DSS))**

I. Introduction

Historically, BI and GIS technologies have followed separate development and implementation paths [10].Decision makers in health sector request for a more complete operational picture and the ability to be more proactive have led to the combination of these two technologies to have suitable tool for Decision Support System (DSS).

Recently, A modern and effective spatial DSS having all the requisite support technologies like online analytical processing (OLAP), specialized analysis and reporting is required for future planners and decision makers; new and efficient methods are needed to integrate the related Information technologies to discover knowledge from large spatial databases [21]

Existing studies focused on spatial and numerical problem solving by using online analytical processing OLAP and GIS system combined to develop the Spatial OLAP Visualization and Analysis Tool (SOVAT) which currently used to solve problems that faces decision maker at health sector as shown in Fig. 1[21].

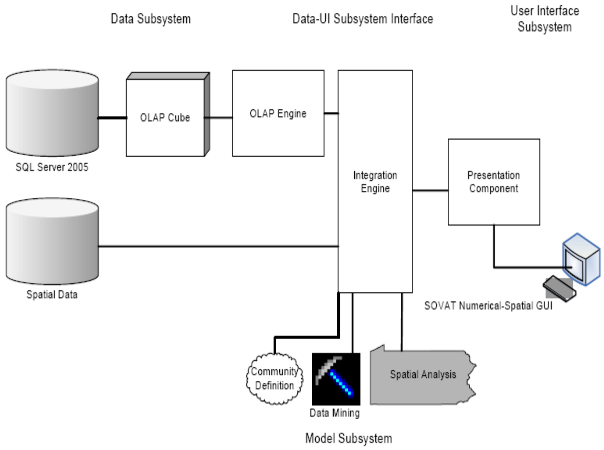


Fig 1 :OlAP-GIS Architecture. Source: University of Pittsburgh 2006

The development of numerical-spatial routines is frequently required to solve complex problems. Individuals who use decision support technology need a system that is capable of supporting the development of numerical-spatial routines integrated with predictive analytical tool.

GIS are designed for the visualization of spatial data and are not perfectly suited for space-time data. That's because today's GIS are built on spatial[18], rather than space-time data structures, GIS maps are static, while the underlying data are dynamic, visualization and the analysis of dynamic marketplace data will benefit greatly from tools that allow the user to animate and interact with maps and graphical data views.

Current geographic knowledge discovery (GKD) methods generally use very simple representations of geographic objects and spatial relationships. Geographic data mining methods should recognize more complex geographic objects (i.e. lines and polygons) and relationships (i.e. non-Euclidean distances, direction, connectivity, and interaction through attributed geographic space such as terrain). Furthermore, the time dimension needs to be more fully integrated into these geographic representations and relationships.

Therefore, having a spatial temporal analysis capability could provide an alternate way to look at public health information. Presenting information with a spatial component triggers a different thought process than seeing the same information on a table or a grid and guide decision maker brainstorm for new possibilities.

Traditional DSS use relational database generators (the middle components that represent the unique software included in the system and are built by the DSS tools) to store the numerical data. GIS is essential for developing spatial routines. There is no alternative technology that could simulate this process.

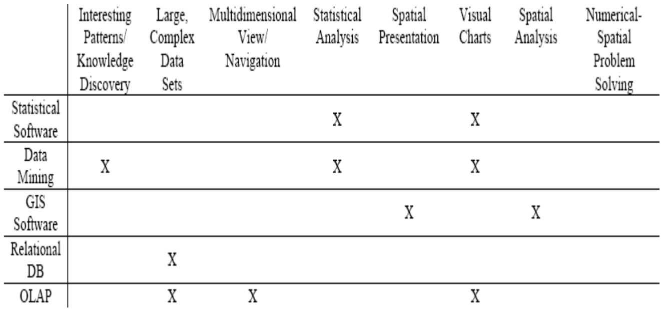
Therefore it must be coupled with a BI that can, by itself, support development of powerful temporal-numerical routines [21]. In addition, it must enable for the development of an interface that supports not only the combination of numerical and spatial information display, but also a temporal interactive easy-to-use environment for creating different types of numerical and spatial routines. As the result, the potential of this synergy is an approach that can significantly enhance spatial, temporal and numerical problem solving through predicative concept

II. Related Work

The increasing availability of (GIS) in health organizations, together with the proliferation of spatially disaggregate data, has led to a number of studies that have been concerned with developing measures of access to health care services. The main aim of this section is to review the use of GIS-based measures in exploring the relationship between geographic access, utilization, quality and health outcomes. There are previous studies in this area. - At first studies about the uses of geography information system in the field of health, for tracking the spread of a disease such as cancer or research area[15]. The second studies in health indicators and tackling health inequalities but don't use spatial predicative tool to support decision makers [7] .

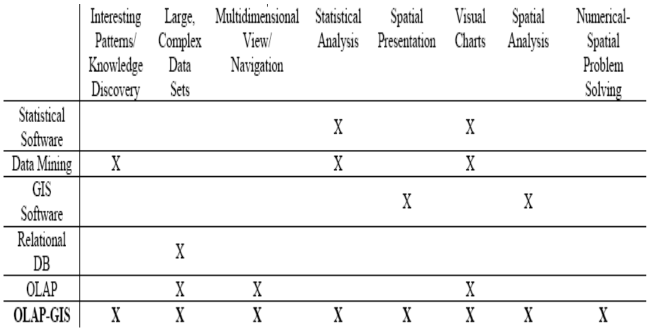
The third studies rely on static spatial tool to make analytical process but not in different time frame [18]. The fourth study an OLAP-GIS technology [2], It is developed at the University of Pittsburgh, The inputs of this research are depend on both OLAP and GIS technologies stand alone as depicted in Table 1 but the outputs of this study is to combine OLAP AND GIS technologies but it has no predicative and temporal concept as shown in Table 2

TABLE 1: DSS GENERATORS FOR DECISION SUPPORT AND CAPABILITIES THEY PROVIDE SOURCE: UNIVERSITY OF PITTSBURGH 2006



	Interesting Patterns/ Knowledge Discovery	Large, Complex Data Sets	Multidimensional View/ Navigation	Statistical Analysis	Spatial Presentation	Visual Charts	Spatial Analysis	Numerical-Spatial Problem Solving
Statistical Software				X		X		
Data Mining	X			X		X		
GIS Software					X		X	
Relational DB		X						
OLAP		X	X			X		

TABLE 2: TRADITIONAL DSS GENERATORS VERSUS AN OLAP-GIS SYSTEM: SOURCE: UNIVERSITY OF PITTSBURGH 2006



	Interesting Patterns/ Knowledge Discovery	Large, Complex Data Sets	Multidimensional View/ Navigation	Statistical Analysis	Spatial Presentation	Visual Charts	Spatial Analysis	Numerical-Spatial Problem Solving
Statistical Software				X		X		
Data Mining	X			X		X		
GIS Software					X		X	
Relational DB		X						
OLAP		X	X			X		
OLAP-GIS	X	X	X	X	X	X	X	X

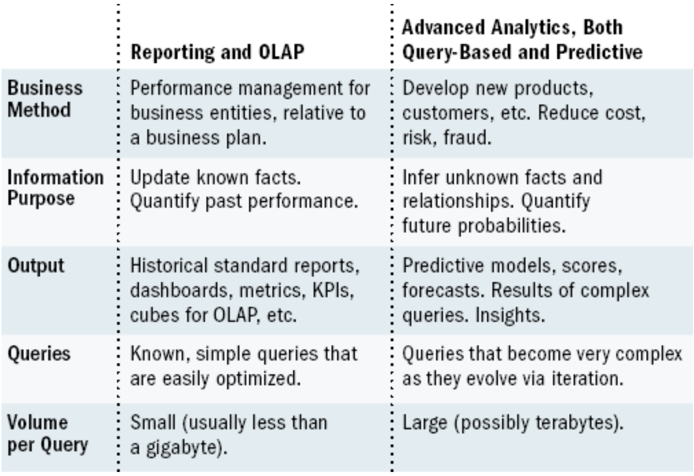
The outputs of this study is using proposed BIGIS_DSS approach to solve health inequality spatial problem which faces decision maker at health sector as shown in Table 4.

But Reporting and analytics are two different practices that have different goals, methods, sponsors, funding, and enabling technologies. Yet many people are being confused between both of them, perhaps because most vendors' platforms for business intelligence (BI) include functions for various types of reporting and summarized analysis in the form of (OLAP)[20].

By comparison, predictive analytics (which includes techniques for data mining and forecasting) is far more exploratory and forward-looking than reporting and (OLAP)[20].The value of predictive analytics is the discovery of unknown facts and relationships, the confirmation of known or suspected relationships, and the leverage of those relationships for better decision making [20].

Achieving these discovery-driven goals through reporting is unlikely, since most reports are based on a small amount of summarized information that's already well known and studied. Likewise, OLAP is usually implemented as a form of parameterized reporting, where the parameters represent dimensions. In such implementations, the available parameters limit the breadth of the analysis, and the analysis cannot be broadened without technical personnel developing more parameters. Table 3 summarizes the differences between reporting and analytics

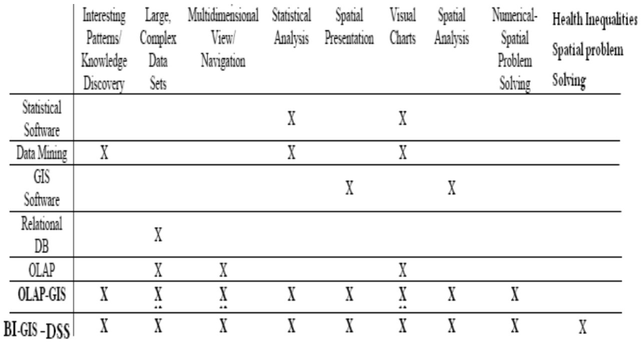
TABLE 3: THE DIFFERENCES BETWEEN REPORTING AND ANALYTICS



	Reporting and OLAP	Advanced Analytics, Both Query-Based and Predictive
Business Method	Performance management for business entities, relative to a business plan.	Develop new products, customers, etc. Reduce cost, risk, fraud.
Information Purpose	Update known facts. Quantify past performance.	Infer unknown facts and relationships. Quantify future probabilities.
Output	Historical standard reports, dashboards, metrics, KPIs, cubes for OLAP, etc.	Predictive models, scores, forecasts. Results of complex queries. Insights.
Queries	Known, simple queries that are easily optimized.	Queries that become very complex as they evolve via iteration.
Volume per Query	Small (usually less than a gigabyte).	Large (possibly terabytes).

/

TABLE 4: BIGIS_DSS PROPOSED APPROACH FOR SOLVING HEALTH INEQUALITY SPATIAL PROBLEM



	Interesting Patterns/ Knowledge Discovery	Large, Complex Data Sets	Multidimensional View/ Navigation	Statistical Analysis	Spatial Presentation	Visual Charts	Spatial Analysis	Numerical-Spatial Problem Solving	Health Inequalities Spatial problem Solving
Statistical Software				X		X			
Data Mining	X			X		X			
GIS Software					X		X		
Relational DB		X							
OLAP		X	X			X			
OLAP-GIS	X	X	X	X	X	X	X	X	
BI-GIS -DSS	X	X	X	X	X	X	X	X	X

Therefore, There is a need for proposed approach to mix GIS,MIS and BI and make them work together to facilitate decision support among varied sections[3].

III. PROPOSED BIGIS_DSS APPROACH

This combination should mend to cover all areas in public health

The main aim of this paper is building a spatial intelligence and predictive analytics approach for solving spatial problem which faces decision makers at health sector as shown in Fig. 2.

-The approach include four phases:

A. *Phase I:*

Build Geodatabase which include spatial data and non spatial data ,spatial data like :

Governorates layer,Streets layer,Rivers layer,

Buildings layer,Usage layer,Borders layer,All of these layers called base map layers

-Non spatial data like

Health data, Health Indicators (as listed below) Health Services, Disease data, Censes information, Health coverage areas information, (We can use database engine like oracle 11g to build this geodatabase).

The following list includes common indicators at health sector:-

-Demographic Indicators

-Health / services Indicators

-Health /Communicable diseases

-Health /non Communicable diseases

- Health Manpower Indicators

-Health Care Indicators

-Nutritional Indicators

-Morbidity Indicators

-Indicators of Resources, Access, and Coverage

-Health facility indicators

-Mortality Indicators

B. *Phase II:*

Build Data Warehouse schema for our geodatabase which enable multi functions like enter query, design charts, analyzing model ,techniques and reporting

C. *Phase III:*

Geoprocessing to convert our data warehouse schema to spatial data warehouse schema using (BI) techniques and spatial model builder tool to build outputs dynamic maps, forecasting and spatial analytical for all common health indicators which effect on health sector .

D. *Phase IV:*

Spatial BIGIS-DSS depend on space-time query to support decision maker at health sector

The aim of this paper is achieved after phase IV implementation for having a spatial intelligence and predictive analytics approach, that'll be used for solving spatial problem which faces decision makers at health sector like health inequalities and non served areas by healthcare.

The proposed approach can help decision makers to solve health problem and restructure the distribution of health units in new areas which outcome from the predicative analytics from the proposed approach and put their expectation for any new health units in future. As well as to predict the spread of diseases and geographical epidemiology epidemics and the spread of disease foci(hotspots) in future .

There is a clear need for an application that is capable of collecting, organizing, and storing temporally and spatially distributed indicator data, generation of maps and other charts. The need for (as listed below) that can eventually be used for the good quality data is an obvious, but unfortunately often overlooked [15], prerequisite to generating high quality thematic maps and analysis from data. It is this gap in quality data collection and organization that proposed standard approach seeks to fill.

Data mining offers great potential benefits for GIS-based applied decision-making. Recently, the task of integrating these two technologies has become of critical importance, especially as various public and private sector organizations possessing huge databases with thematic and geographically referenced data begin to realize the huge potential of the information contained therein.

Challenges in spatial mining: Geospatial data repositories tend to be very large. Moreover, existing GIS datasets are often splintered into feature and attribute components that are conventionally archived in hybrid data management systems. Algorithmic requirements differ substantially for relational (attribute) data management and for topological (feature) data management. [35]

Related to this is the range and diversity of geographic data formats, which present unique challenges. The digital geographic data revolution is creating new types of data formats beyond the traditional "vector" and "raster" formats. Geographic data repositories increasingly include ill-structured data, such as imagery and geo-referenced multi-media [22].

IV. BIGIS_DSS Approach Features

As mentioned earlier in this paper eventually lead to a better understanding of the relationship between the adoption of BI,GIS,DSS process and its impact in a health sector environments. This section discusses the BIGIS_DSS Approach Features.

Spatial data mining is the application of data mining methods to spatial data. The end objective of spatial data mining is to find patterns in data with respect to geography. So far, data mining and Geographic Information Systems (GIS) have existed as two separate technologies, each with its own methods, traditions, and approaches to visualization and data analysis. Particularly, most contemporary GIS have only very basic spatial analysis functionality. The immense explosion in geographically referenced data occasioned by developments in IT, digital mapping, remote sensing, and the global diffusion of GIS emphasize the importance of developing data-driven inductive approaches to geographical analysis and modeling.

The paper's aim is exploratory, building towards understanding and subsequently towards constructive guidance for the adoption process of new technologies in such as BI, GIS and motion charts, sometimes in hospitable, health environments. Since the paper deals with poorly understood issues in a rich, difficult to control environment, caution ought to be taken in assessing the results. However, taking the exploratory design's restraints into account, two theoretical contributions can be derived:

1-Drawing attention to identified potential new technology adoption approach such as BI,GIS and DSS that determine spatial intelligence and predictive analytics for solving spatial problem which faces decision makers at health sector .

2- Discussing how the successful adoption of a new approach increases its impact in health sector environments.

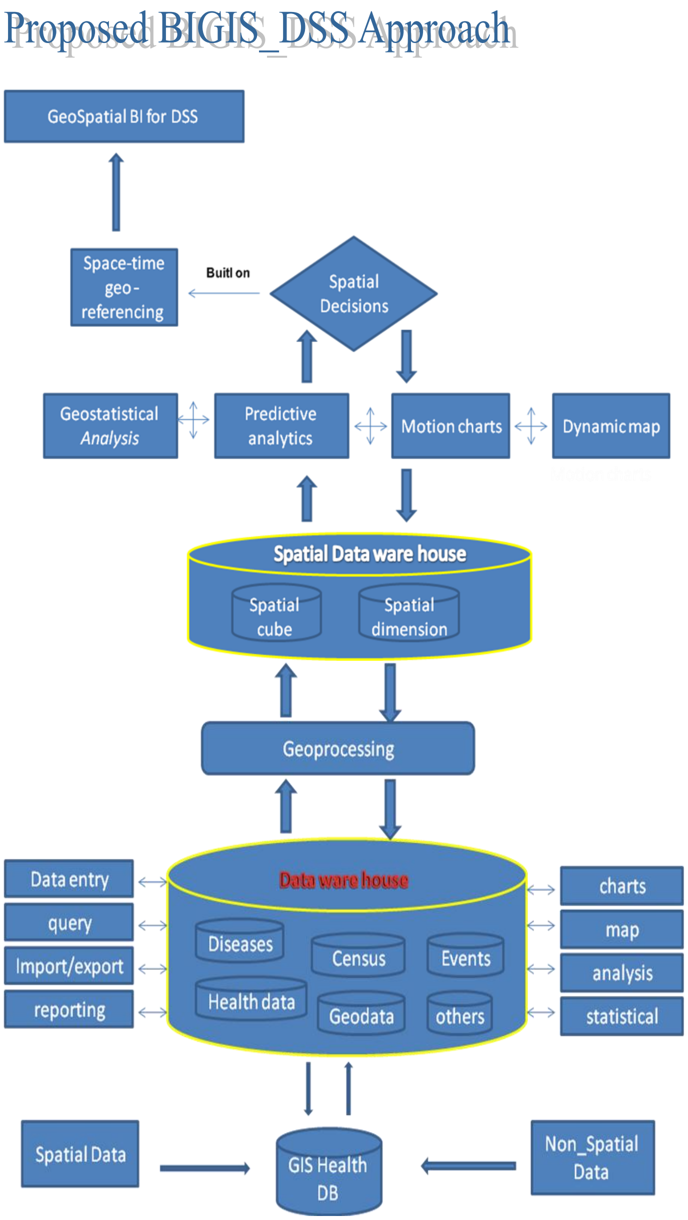


Fig. 2: The Proposed BIGIS-DSS Approach at Health Sector

As wide as the possible applicability of the study is, the claim to managerial contribution should be just as modest. An exploratory study cannot provide normative [18]. The reason for this is that case studies in principle can only establish the existence of relations but not their direction [19]. In other words, it is only possible to conclude that one variable is related to another variable, but impossible to determine the dependent and independent one. In addition, generalizing from a single case study however carefully selected cannot be justified [20] Nonetheless given these constraints, this paper will be useful to the following stakeholders:

1-Senior managers of the health departments who develop and improve health services should be interested in the results, because the results deal with a management issue of strategic importance database which found in BI&GIS servers as in Fig. 3.

Therefore this paper will combine a number of international health indicators and indicators associated spatial analysis and the work necessary to extract the proposed standard approach criteria for health indicators can be used in the spatial scope of local and international health.

This paper expects that combining the strength in predicate analysis and the ease of use makes the proposed approach ideal for healthcare professionals without extensive computer skills. And also expects that the maps resulting from the proposed approach will provide decision makers with information to strengthen their disaster management capacity. It also will represent the basis for the reflection that needs to take place regarding populations' vulnerability towards natural hazards from a health perspective as shown in Figure 4

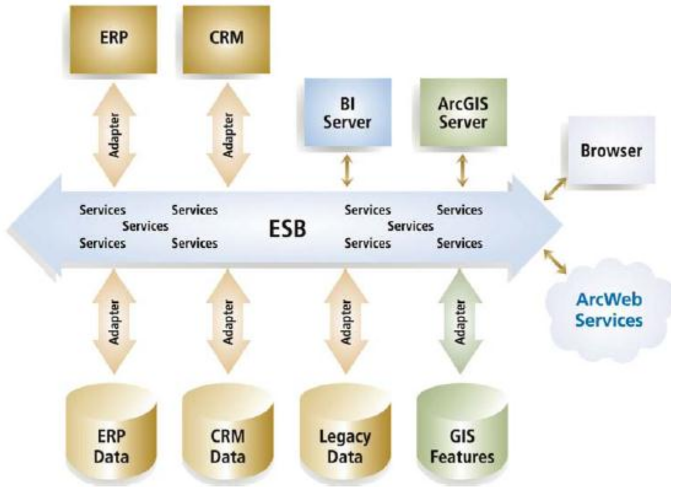


Fig 3:

A simplified view of service-oriented architecture utilizing the enterprise service bus (ESB) to pass services between various applications. Applications in ArcGIS Server can use aggregated data managed by the BI server and utilize its reporting platform. The ArcGIS Server could also use enterprise resource planning (ERP) or CRM data via a service independent of the BI server.

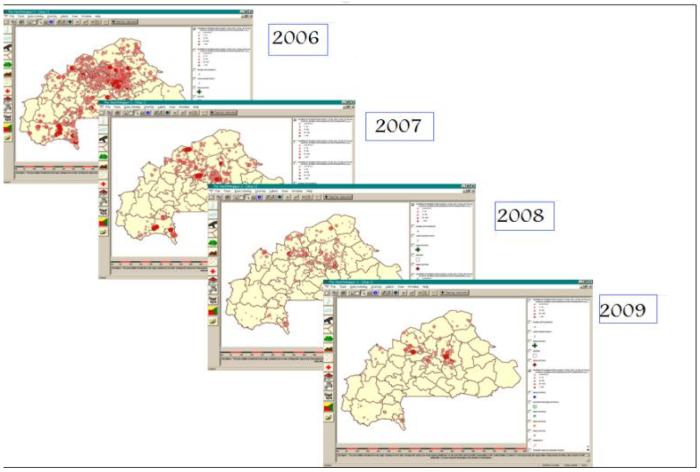


Fig. 4: Comparison between Health Maps in Different Years

Source :WHO site [information system department],2009

The approach covers the following items:

1-Reviewing that there are a variety of health indicators, according to several different destinations of health sector.

2-Attempt to provide a set of indicators to cover the full breadth of the approach. Identification of the spatial location of a geographic community by using GIS, Identification of health factors within the community by using numerical data such as death counts, disease incidence or prevalence rates, Identification of the spatial location of bordering communities of interest using GIS.

3- Development of a dynamic proposed spatial approach for health information and indicators

4-Implementing the proposed approach by inserting the health information and comparing subset of health indicators before and after using the proposed spatial approach of health indicators to display the difference and how the proposed approach can help in tackling health inequalities

5-Reviewing that spatial approach of health indicators will enhance presentation of the health inequalities dimensions at the health sector .

6-Using spatial approach of health indicators for DSS to improve health sector.

V. The Proposed Methodlogy

-This paper depend on the following methodology to implement the proposed approach :-

Phase I : Data collection phase (spatial data and row data)

Phase II : Building geodatabase through different software integration and insert data to database

Phase III : Developing proposed approach by using different integration tools

Phase IV : Data analyzing and processing depend on spatial statistical analyze tool for geodatabase of health indicators.

Phase V : Test the output of information through development proposed spatial analytical approach for health information indicators

Phase VI : Produce dynamic health map for DSS and follow up phase through implementation of DSS process for health indicators to restructure all resources, as shown in Figure 5.

VI. Site Selection

Identifying a new service location is one of the most basic functions of business development. Having the ability to quickly access the geodemographic dynamics of health sector in contrast to the likely demand for services at a new location requires flexible powerful analytical software tools.

ArcGIS desktop, Business Analyst, ArcGIS server ,geoprocessing , OLAP and BusinessMAP PRO software all provide varying levels of capabilities for site selection and combination between all of them is suitable for initial design for proposed approach as shown in Figure 5.

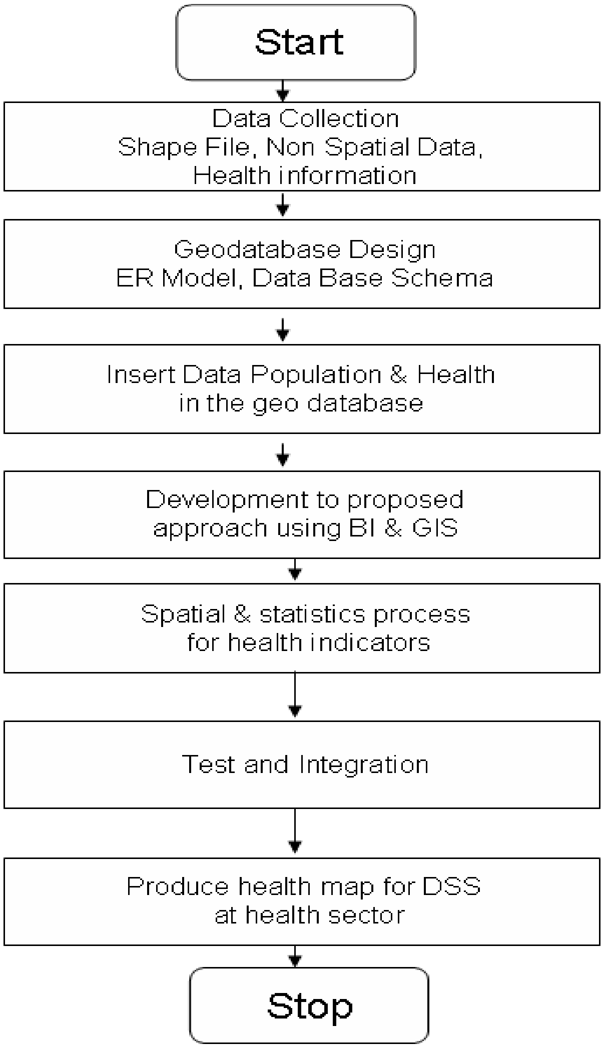


Fig. 5 : Methodology Phases

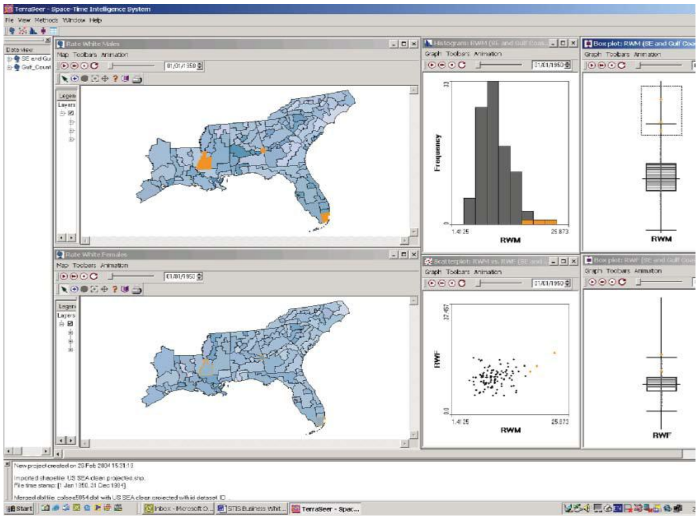

TerraSeer - Space-Time Intelligence System
File View Methods Window Help
Frequency
RWM
RWF

Fig. 6：The Proposed Approach with Different software

VII. Conclusion And Future Work

BIGIS-DSS proposed approach enhances OLAP-GIS at health sector by making adoption of BI,GIS and DSS technologies to build a spatial intelligence and predictive analytics approach for solving spatial problem which faces decision makers at health sector. The proposed spatial analytical approach will cover three main health planning issues. These issues are tackling health inequalities through geospatial monitor for inequalities in distribution of health units and its services, support decision-making with predictive analytics for common health indicators, and geoprocessing for input layers through dynamic health map and motion charts to support decision making. This paper shows that combining is the strength in predicate analysis and the ease of use makes the proposed approach ideal for healthcare professionals without extensive computer skills. The maps resulting from the proposed approach will provide decision makers with information to strengthen their disaster management capacity. It also will represent the basis for the reflection that needs to take place regarding populations' vulnerability towards natural hazards from a health perspective. BIGIS-DSS approach is able to be integrated and cooperated with other sciences through different types of information systems and, it can be used as a new direction that will be deployed for both national and international development business. The proposed approach can be considered as a nucleus to increase the economy of developing countries at health sector.

The results demonstrate the potential for BIGIS_DSS in health sector analysis. Future work will explore the impact of the BIGIS_DSS system in other areas of public health as using BIGIS_DSS to develop slums in development countries.

References

[1] Anton F, Oldfield E, Coleman DJ,Towards web-based representation aand processing of health information Gao S, Mioc D, Yi X, Innternational Journal of Health Geographics 2009, 8:3

[2] Ben,Eazzetts ,GIS:The Next generation The Convergence of Technologies ,map Middle east,USA 2009

[3] Etienne Dubé,Thierry Badard ,An Introduction to geomonderian and spatialytics, Laval University, Quebec, Canada,2009

[4] Geoffrey M Jacquez and Robert Rommel , Local indicators of geocoding accuracy (LIGA): theory and application, International Journal of Health Geographics 2009, 8:60doi:10.1186/1476-072X-8-60

[5] Goldberg D: A Geocoding Best Practices Guide. Springfield, IL North American Association of Central Cancer Registries, 2008.

[6] Health indicators ,http://www.mohp.gov.eg/sec/About/engaza/2.doc

[7] Henry KA, Boscoe FP: Estimating the accuracy of geographical imputation .International Journal of Health Geographics 2008, 7:3.

[8] Kravets N, Hadden W: The accuracy of address coding and the effects of coding errors. Health Place 2007, 13:293-298.

[9] Matthew, Laurence Scotch, AN OLAP-GIS SYSTEM FOR NUMERICAL-SPATIAL PROBLEM SOLVING IN COMMUNITY HEALTH ASSESSMENT ANALYSIS, University of Pittsburgh ,2006

[10] Monaco Va1,parmanto Bambang, Scotch Matthew, Evaluation of SOVAT: An OLAP-GIS Decision Support System for Community Health Assessment Data Analysis, BMC Medical Informatics and Decision Making, Year :2008,issue 1 ,pp:10,Vol:8

[11] Philip ,Russom, Data Requirements for Advanced Analytics, TDWI Checklist report,2008

[12] Refaat M Kamell and Samir M Wassif, Population Problem In Egypt As One of The Priority Health Problems In Tropical Areas Surgery

[13] Department, Faculty of Medicine, Ain-Shams University and Community Medicine Department, Faculty of Medicine, Zagazig University, Egypt_2008

[14] Rushton G, Armstrong M, Gittler J, Greene B, Pavlik C, West M, Zimmerman D. Boca Raton, Statistical methods for incompletely and incorrectly geocoded cancer data. In Geocoding Health Data. PL: CRC Press, 2008

[15] Space-Time Intelligence System to improve retail decision making, http://www.terraseer.com/products_stis.php

[16] Terban , Efraim., Decision Support and Business Intelligence Systems [et al.].8th ed.p.cm.2007

[17] Ward MH, Nuckols JR, Giglierano J, Bonner MR, Wolter C, Airola M, Mix W, Colt JS, Hartge P: Positional accuracy of two methods of geocoding. Epidemiology 2005, 16:542-547

[18] Wassif S.M and Amany R.: Spot lights on the demographic profile of Egypt, The Egyptian journal of Community Medicine 2008Vol. 26, No.1

[19] Wei Gu , Xin Wang, S Elizabeth McGregor, Optimization of preventive health care facility locations, Canada,2010

[20] Zahran M.: Evaluation of census,2006. Al Ahram newspaper, 4/4/2007.

[21] Zandbergen PA: Positional Accuracy of Spatial Data: Non-Normal Distributions and a Critique of the National Standard for Spatial Data Accuracy Transactions in GIS 2008, 12:103-130.

[22] Zimmerman D, Fang X, Mazumdar S, Rushton G: Modeling the probabilitydistribution of positional errors incurred by residential address geocoding International Journal of Health Geographics 2007, 6:1. Booman M, et al.: Using a geographical information system to plan a malaria control programme in South Africa.Bull World Health Organ 2000, 78:1438-1444.

[23] Peters DH, Garg A, Bloom G, Walker DG, Brieger WR, Rahman MH: Poverty and access to health care in developing countries.

[24] Annals of the New York Academy of Sciences 2008, 161-171.1136(Reducing the Impact of Poverty on Health and Human Development: Scientific Approaches)

[25] Müller I, Smith T, Mellor S, Rare L, Genton B: The effect of distance from home on attendance at a small rural health centre in Papua New Guinea.International Journal of Epidemiology 1998, 27(5):878.

[26] Arcury TA, Gesler WM, Preisser JS, Sherman J, Spencer J, Perin J: The effects of geography and spatial behavior on health care utilization among the residents of a rural region.Health Services Research 2005, 40(1):135-156.

[27] Tanser F, Gijsbertsen B, Herbst K: Modelling and understanding primary health care accessibility and utilization in rural South Africa: An exploration using a geographical information system.Social Science & Medicine 2006, 63(3):691-705.

[28] Tanser F, Wilkinson D: Spatial implications of the tuberculosis DOTS strategy in rural South Africa: a novel application of geographical information system and global positioning system technologies.Tropical Medicine & International Health 1999, 4(10):634-638.

[29] Perry B, Gesler W: Physical access to primary health care in Andean Bolivia.Social Science & Medicine 2000, 50(9):1177-1188.

[30] McLafferty SL: GIS and health care, 2003.

Personalized Semantic Retrieval and Summarization of Web Based Documents

Salah T. Babekr
Computer Science Dept., College of Computers and Information Technology, Taif University, Kingdom of Saudi Arabia (KSA)

Khaled M. Fouad
Computer Science Dept., College of Computers and Information Technology, Taif University, Kingdom of Saudi Arabia (KSA)

Naveed Arshad
Computer Science Dept., College of Computers and Information Technology, Taif University,Kingdom of Saudi Arabia (KSA)

***Abstract* —The current retrieval methods are essentially based on the string-matching approach lacking of semantic information and can't understand the user's query intent and interest very well. These methods do regard as the personalization of the users. Semantic retrieval techniques are performed by interpreting the semantic of keywords. Using the text summarization allows a user to get a sense of the content of a full-text, or to know its information content, without reading all sentences within the full-text.**

In this paper, a semantic personalized information retrieval (IR) system is proposed, oriented to the exploitation of Semantic Web technology and WordNet ontology to support semantic IR capabilities in Web documents. In a proposed system, the Web documents are represented in concept vector model using WordNet. Personalization is used in a proposed system by building user model (UM). Text summarization in a proposed system is based on extracting the most relevant sentences from the original document to form a summary using WordNet.

The examination of the proposed system is performed by using three experiments that are based on relevance based evaluation. The results of the experiment shows that the proposed system, which is based on Semantic Web technology, can improve the accuracy and effectiveness for retrieving relevant Web documents.

Keywords-Semnatic Web; WordNet; Personalization; User Model; Information Retrieval; Summerization.

I. INTRODUCTION

Internet access, such as World Wide Web (WWW), has made document retrieval increasingly demanding as collection and searching of documents has become an integral part of many people's lives. Accuracy and speed are two key measurements of effective retrieval methodologies. Existing document retrieval systems use statistical methods [1] and natural language processing (NLP) [2] approaches combined with different document representation and query structures. Document retrieval [3] has created many interests in the information retrieval (IR) community.

Document retrieval refers to finding similar documents for a given user's query. A user's query can be ranged from a full description of a document to a few keywords. Most of the extensively used retrieval approaches are keywords based searching methods, e.g., www.google.com, in which untrained users provide a few keywords to the search engine finding the relevant documents in a returned list [4]. Another type of document retrieval is to use a query context by using language modeling, to integrate several contextual factors so that document ranking will be adapted to the specific query contexts [5]. Using an entire document as a query performs well in improving retrieval accuracy, but it is more computationally demanding compared with the keywords based method [6].

The effectiveness of processes models based on keywords is limited by the phenomenon known as "keywords barrier", i.e., the internal representation of an information item by a set of words extracted from texts through statistical and / or syntactic techniques does not allow a considerable improvement of the effectiveness of IR systems and, in particular, the precision of their results. These limitations have stimulated the development of several techniques trying to extract meaning from texts, such as semantic analysis [7] to obtain more accurate internal representations of information items. However, there is a lack of semantic retrieval process models providing appropriate abstraction representations of the activities, products and techniques involved in such retrieval processes [8].

Several IR process models, such the Boolean [9], the vector space [10] and the probabilistic models [11] have been proposed to cover the activities and technical user queries as well as storage and retrieval of information items from unstructured sources. Classic models represent the documents with a set of keywords extracted from text and propose different approaches to retrieval and presentation of retrieved information items sorted according to their relevance.

Some of the reasons that the classical IR approaches tend to be less effective as the web evolves can be identified as follows:

Content of the current web is created using natural language and HTML is a formatting language which is used to render presentation to human. The content of the web pages are not understandable with agents.

- Classical IR models are based on the computation of words or word occurrence which is a semantically imprecise calculus.

- The metadata is not available with the current web resources and there is no such a standard for creating the metadata.
- Interoperability and reusability of the web content is difficult due to heterogeneity of the web contents.

Personalized Semantic retrieval and summarization architecture aims at improving the conventional IR which is based on semantic Web technology. The personalized semantic enhanced retrieval and summarization framework is proposed that meets our objectives. The work begins with an overview of the research and then provides a comprehensive literature review on the related research topics. In particular, we conducted a selected study on the existing semantic IR systems and provide a detailed survey. More importantly, we suggest some improvements after the study of the existing systems. The idea also outlines our methodology towards designing a personalized semantic IR system.

II. Basic Concepts

A. *Semantic Web*

The Semantic Web [12] is a Web-based technology that extends XML by providing the means to define ontologies; the definition of objects and relationships between them. This allows machines to make intelligent inferences about objects across the Web. This allows intelligent agents [13] embodying knowledge about certain aspects of software development (much of it may be organization-specific) to make intelligent inferences that can be used as the basis for improved decision-making on software development processes, and usability issues. In addition, the semantic web is an approach to facilitate communication by making the web suitable for machine-to-machine communication [14]. It can be used to encode meaning and complex relationships in web pages. A major challenge for the emerging semantic-web field is to capture the knowledge required and structure it in a format that can be processed automatically (e.g., by agents).

Informally, ontology [15] of a certain domain is about terminology (domain vocabulary), all essential concepts in the domain, their classification, their taxonomy, their relations (including all important hierarchies and constraints), and domain axioms. More formally, to someone who wants to discuss about topics in a domain using a language, ontology provides a catalogue of the types of things assumed to exist in a domain; the types in the ontology are represented in terms of the concepts, relations, and predicates of language.

WordNet

WordNet [16, 17] has been used in several capacities to improve the performance of IR systems. WordNet can be used to solve the research problems in IR.

To overcome the weaknesses of term-based representation that is found in the conventional IR approaches, an ontology-based representation has been recently proposed [18], which exploits the hierarchical is-a relation among concepts, i.e., the meanings of words. For example, to describe with a term-based representation documents containing the three words: “animal”, “dog”, and “cat” a vector of three elements is needed; with an ontology-based representation, since “animal” subsumes both “dog” and “cat”, it is possible to use a vector with only two elements, related to the “dog” and “cat” concepts, that can also implicitly contain the information given by the presence of the “animal” concept. Moreover, by defining an ontology base, which is a set of independent concepts that covers the whole ontology, an ontology-based representation allows the system to use fixed-size document vectors, consisting of one component per base concept.

In the text representation, the terms are replaced by their associated concepts in WordNet [19]. In the pretreatment phase, it firstly convert uppercase characters into lowercase characters and then eliminate from text punctuation marks and stop words such as: are, that, what, do. This representation requires two more stages: a) the “mapping” of terms into concepts and the choice of the “merging” strategy, and b) the application of a disambiguation strategy. The first stage is shown in example, as found in figure 1, is about mapping the two terms government and politics into the concept government (the frequencies of these two terms are thus cumulated). Then, among the three “merging” strategies offered by the conceptual approach (“To add concept”, “To replace terms by concepts” and “concept only”), the strategy “concept only“ can be chosen, where the vector of terms is replaced by the corresponding vector of concepts (excluding the terms which do not appear in WordNet).

Voorhees [20] suggested that WordNet can be used in IR for query expansion. Query expansion is considered to be one of the techniques that can be used to improve the retrieval performance of short queries. Most of the indexing and retrieval methods are based on statistical methods; short queries posed challenges to this model due to the limited amount of information that can be gathered during its processing. In expanding the query, Voorhees suggested using of synonyms, hypernyms, hyponyms, and their combinations. The results showed that using of synonyms, hypernyms, and hyponyms are significant in the retrieval performance for short queries, but little improvement when they are applied to the long query.

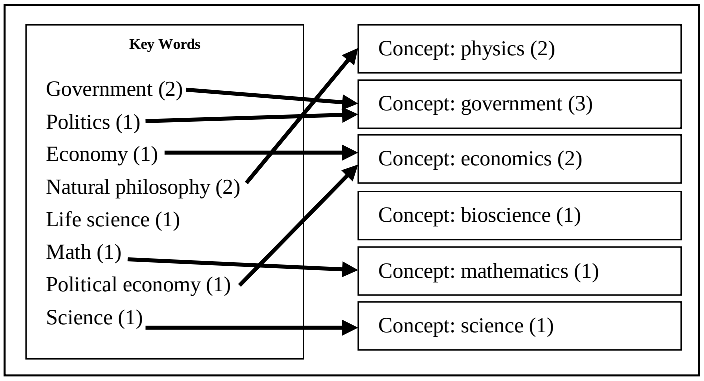


Figure 1. Example of mapping words in concepts

B. *Personalization*

The goal of personalization [21] is to endow software systems with the capability to change (adapt) aspects of their functionality, appearance or both at runtime to the particularities of users to better suit their needs. The recent rapid advances in storage and communication technologies stress the need for personalization. This need is more evident in consumer oriented fields, like news content personalization systems, recommendation systems, user interfaces, and applications like home audiovisual material collection and organization, search engines in multimedia browsing and retrieval systems, providing services for personalized presentation of interactive video. The core idea of personalization is to customize the presentation of information specifically to the user to make user interfaces more intuitive and easier to understand, and to reduce information overload.

User modeling [22] describes the process of creating a set of system assumptions about all aspects of the user, which are relevant to the adaptation of the current user interactions. This can include user goals, interests, level of expertise, abilities and preferences. The most reliable method of user modeling is by explicit entry of information by the user. In most practical systems, this is too time-consuming and complex for the user. Hence implicit user modeling, based on analysis of past and current user interactions, is critical. The user profile is a machine-processable description of the user model [23, 24].

C. Text summarization

Text summarization [25] is a data reduction process. The use of text summarization allows a user to get a sense of the content of a full-text, or to know its information content, without reading all sentences within the full-text. Data reduction increases scale by (1) allowing users to find relevant full-text sources more quickly, and (2) assimilating only essential information from many texts with reduced effort. Text summarization is particularly useful in certain domain, where oncologists must continuously find trial study information related to their specialty, evaluate the study for its strength, and then possibly incorporate the new study information.

Text Summarization [26] methods can be classified into extractive and abstractive summarization. An extractive summarization method consists of selecting important sentences, paragraphs etc. from the original document and concatenating them into shorter form. The importance of sentences is decided based on statistical and linguistic features of sentences. An Abstractive summarization [26] attempts to develop an understanding of the main concepts in a document and then express those concepts in clear natural language.

Extractive summaries [26, 27]are formulated by extracting key text segments (sentences or passages) from the text, based on statistical analysis of individual or mixed surface level features such as word/phrase frequency, location or cue words to locate the sentences to be extracted.

Extractive text summarization process can be divided into two steps [28]:

1) *Preprocessing step*, in this step Sentences boundary identification, Stop-Word Elimination and Stemming are performed and,

2) *Processing step*, in this step features influencing the relevance of sentences are decided and calculated and then weights are assigned to these features using weight learning method. Final score of each sentence is determined using Feature-weight equation. Top ranked sentences are selected for final summary.

III. Related Work

Personalized search [29] is addressed by a number of systems. Persona [30] used explicit relevant feedback to update user profiles that are represented by means of weighted open directory project taxonomy [31]. These profiles are used to filter search results. Personalized variants of PageRank, is as found in PersonalizedGoogle or the Outride Personalized Search System [32]. Persival [33] re-ranked the search results of queries for medical articles profiles keywords, associated concepts, and weights generated from an electronic patient record.

In [34] it was filtered search results on the grounds of user profiles obtained from earlier queries. These profiles consist of a set of categories, and weighted terms associated with each category. In their work on personalizing search results, [35] distinguish between long-term and short-term interests. While aiming at personalization in a broader sense, [36] use click-through data to increase the performance of search results.

Nowadays [37], personalization systems are developed by considering ontology to reduce the limitation of traditional IR such as information overload or cold start problem. So considering ontology to build an accurate profile brings some extra benefit in user modeling. A user profile can be presented as a weighted concept hierarchy for searching and browsing in the web. User profile can be created by user with his/her personal information and interest or it can be a reference one. However, profile can be created by manually entering the user’s information or automatically by watching the use’s activities.

Jin and others [38], proposed a novel approach which enables intelligent semantic web search for best satisfying users search intensions. The proposed approach combines the user’s subjective weighting importance over multiple search properties together with fuzziness to represent search requirements. A special ranking mechanism based on the above weighed fuzzy query is also presented. The ranking method considers not only fuzzy predicates in the query, but also the user’s personalized interests or preferences.

MedSearch is a complete retrieval system for medical literature [39]. It supports retrieval by SSRM (Semantic Similarity Retrieval Model), a novel IR method which is capable for associating documents containing semantically similar (but not necessarily lexically similar) terms. SSRM suggests discovering semantically similar terms in documents and queries using term taxonomies (ontologies) and by associating such terms using semantic similarity methods. SSRM demonstrated very promising performance achieving significantly better precision and recall than Vector Space Model (VSM) for retrievals on Medline.

In [40], the authors proposed a new approach to User Model Acquisition (UMA) which has two important features. It doesn’t assume that users always have a well-defined idea of what they are looking for, and it is ontology-based, i.e., it was dealt with concepts instead of keywords to formulate queries.

The first problem is that most approaches assume users to have a well-defined idea of what they are looking for, which is not always the case. They solved this problem by letting fuzzy user models evolve on the basis of a rating induced by user behavior. The second problem concerns the use of keywords, not concepts, to formulate queries. Considering words and not the concepts behind them often leads to a loss in terms of the quantity and quality of information retrieved. They solved this problem by adopting an ontology-based approach.

In [41], authors introduced a method for learning and updating a user profile automatically. The proposed method belongs to implicit techniques. It processes and analyzes behavioral patterns of user activities on the web, and modifies a user profile based on extracted information from user's web-logs. The method relies on analysis of web-logs for discovering concepts and items representing user's current and new interests. The mechanism used for identifying relevant items is built based on a newly introduced concept of ontology-based semantic similarity.

Dıaz and Gervas [42] have proposed the personalized summarization as a process of summarization that preserves the specific information that is relevant for a given user profile, rather than information that truly summarizes the content of the news item. The potential of summary personalization is high, because a summary that would be useless to decide the relevance of a document if summarized in a generic manner, may be useful if the right sentences are selected that match the user interest. Authors defend the use of a personalized summarization facility to maximize the density of relevance of selections sent by a personalized information system to a given user.

Lv, Zheng and Zhang [43] have developed the method of IR based on semantics. In addition, they took the "wine" ontology instances provided by Stanford University as a reference, and develop a Chinese "wine" model by using protégé tools. Finally, the retrieval results show that the proposed method has higher recall and precision.

Rinaldi [44] have given the solution for the problem of IR on the Web using an approach based on a measure of semantic relatedness applied to evaluate the relevance of a document with respect to a query in a given context: the concepts of lexical chains, ontologies, and semantic networks. The proposed methods, metrics, and techniques are implemented in a system called DySE (Dynamic Semantic Engine). DySE implements a context-driven approach in which the keywords are processed in the context of the information in which they are retrieved, in order to solve semantic ambiguity and to give a more accurate retrieval based on the real of the user interests.

Huang and Zhang [45] proposed the approach to expand the set of query keywords based on associational semantics. Firstly, they constructed a group of semantic trees for original keywords one by one based on WordNet, an online lexical system. The original keywords perch on the roots of the trees. Secondly, they removed noise nodes in the trees by computing the similarity between words, and assemble the trees into a big integrated tree, i.e. Tree of Associational Semantics Model, by expanding the roots of the trees upward until finding the common origin of the trees. They assigned a weight to each word on the trees, and selected candidates from the trees by referring to thresholds. Finally, they executed the document retrieval by importing the weights and distribution density of keywords into calculation of similarities between query and documents.

Gauch, Speretta and Pretschner [46] explored the use of ontology-based user profiles to provide personalized search results. In this work, authors used the ontology that consists of hierarchies of concepts in which each concept is defined by a set of documents, and hierarchy is induced by an informal specialization relationship. They reviewed a variety of sources of information from which the ontology-based profiles can be created, and described improvements in accuracy achieved when the user profiles are used to select search results.

IV. Architecture of Personalized Semantic Retrieval and Summarization

The semantic retrieval approach embeds background knowledge with explicitly defined semantics can help to build intelligent IR applications. Based on some of the weaknesses of conventional IR techniques, the motivations towards a semantic IR framework have been identified.

Using of text summarization allows a user to get a sense of the content of a full-text, or to know its information content, without reading all sentences within the full-text. Data reduction increases scale by (1) allowing users to find relevant full-text sources more quickly, and (2) assimilating only essential information from many texts with reduced effort. Text summarization is particularly useful in certain domain, where oncologists must continuously find trial study information related to their specialty, evaluate the study for its strength, and then possibly incorporate the new study information.

The improved and practical approach is presented to automatically summarizing Web documents by extracting the most relevant sentences from the original document to create a summarization. The idea of proposed approach is to find out key sentences from the Keyword extraction based on statistics and synsets extraction using WordNet. These two properties can be combined and tuned for ranking and extracting sentences to generate a list of candidates of key sentences. Then semantic similarity analysis is conducted between candidates of key sentences to reduce the redundancy. The entire architecture of the proposed approach is shown in the figure 2.

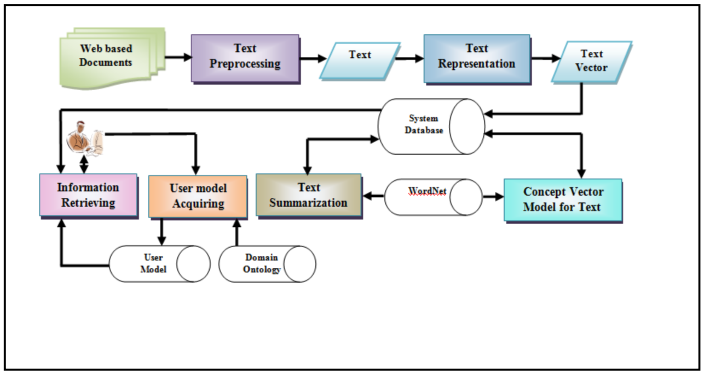


Figure 1. Architecture of the proposed approach

The detail of each process is found in the next sections.

A. Text Preprocessing

The most widely accepted document representation model in text classification is probably Vector Space Model (VSM) [10]. VSM is adapted in the proposed system to achieve effective representations of documents. The documents must be preprocessed before the text representation. The main procedures of preprocessing are shown in figure 3.

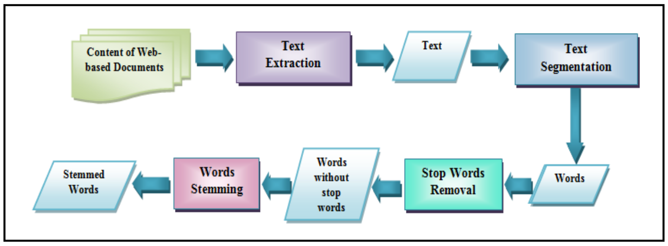


Figure 2. Main steps for text preprocessing

Text Extraction

The first step of the text representation process is extracting textual data from the web pages. Then convert each page into individual text document to apply text preprocessing techniques on it. This step is applied on input Web documents dataset by scanning the web pages and categorizing the HTML tags in each page.

Then exclude the tags that contain no textual information like formatting tags and imaging tags (i.e. <HTML>, <BODY>, <IMG>, etc.). Also exclude all the scripts and codes that are found in the page like JavaScript and VBScript. Then extract the textual data from other tags (like paragraphs, hyperlinks, and metadata tags) and store it into individual text documents as input for next steps. To extract the text from Web documents, open source high-performance .NET C# module is used that was created to parse HTML [47] for links, indexing and other purposes.

1) Stop Words Removal

Stop words, i.e. words thought not to convey any meaning, are removed from the text. In this work, the proposed approach uses a static list of stop words about all tokens. This process removes all words that are not nouns, verbs or adjectives. For example, stop words removal process will remove all the words like: he, all, his, from, is, an, of, your, and so on. Removing these words will save spaces for storing document contents and reduce time taken during the search process.

2) Words Stemming

The stem is the common root-form of the words with the same meaning appear in various morphological forms (e.g. player, played, plays from stem play). In the proposed approach, the morphology function [48] based on WordNet [49] to perform stemming process. Stemming will find the stems of the output terms to enhance term frequency counting process because terms like "computers" and "engineering" come down from the same stem "computer" and " engineer". This process will output all the stems of extracted terms [50, 51].

B. Text Representation

VSM [10] is adapted in the proposed system to achieve effective representations of documents. Each document is identified by n-dimensional feature vector where each dimension corresponds to a distinct term. Each term in a given document vector has an associated weight.

The weight is a function of the term frequency, collection frequency and normalization factors. Different weighting approaches may be applied by varying this function. Hence, a document j is represented by the document vector $\mathrm{d_j}$:

$$\mathrm{d_j} = (\mathrm{w_{1j}}, \mathrm{w_{2j}}, \mathrm{w_{nj}}) \qquad (1)$$

Where, $\mathrm{w_{nj}}$ is the weight of the kth term in the document j.

The term frequency reflects the importance of term k within a particular document j. The weighting factor may be global or local. The global weighting factor clarifies the importance of a term k within the entire collection of documents, whereas a local weighting factor considers the given document only. The document keywords were extracted by using a term-frequency and inverse-document-frequency (*tf-idf*) calculation [52], which is a well-established technique in IR. The weight of term k in document j is represented as:

$$\mathrm{w_{kj}} = \mathrm{tf_{kj}} \times (\log_2^{\mathrm{n}} - \log_2^{\mathrm{df_k}} + 1) \qquad (2)$$

Where: $\mathrm{tf_{kj}}$ = the term k frequency in document j, $\mathrm{df_k}$ = number of documents in which term k occurs, n = total number of documents in collection. The output of this step is the weight of terms in selected document.

C. Concept Vector Model of Text using WordNet

The purpose of this step is to identify WordNet concepts that correspond to document words. Concept identification [53] is based on the overlap of the local context of the analyzed word with every corresponding WordNet entry. The words mapping into concepts algorithm for the terms is given in figure 4.

```
Input: ($B_w$) Bag of words ($W_i$) in document D that was gotten from Words Stemming phase.
Output: Set of all WordNet concepts belonging to terms (words) in document D.
Procedure:
// ($C_w$) is the count of words in the bag, and ($Cont_i$) the context of the word in the document, it is the sentence in document D that contains the word occurrence being analyzed.
Do While $i <= C_w$
        Get WordNet entries $C_i$ set ($\mathrm{CSet_i}$) that is containing the word $\mathrm{W_i}$ , where $C_i \in \mathrm{CSet}_i$.
        Save $\mathrm{W_i}$ and its $C_i$ in database table.
EndDo
  Rank concepts $\mathrm{C_i}$ in $\mathrm{CSet_i}$ where $|\mathrm{C_1}| > |\mathrm{C_2}| > |\mathrm{C_3}| \ldots > |\mathrm{C_n}|$ // |  | denotes the concept length, in terms of the number of words in the corresponding terms. $CSet_{ri}$ is the ranked concepts set.
FOR each $\mathrm{C_i}$ in $\mathrm{CSet_{ri}}$
          Get common words between $\mathrm{Cont_i}$ and representative term of $\mathrm{C_i}$ ,
          which is the intersection $\mathrm{Cint} = \cap (\mathrm{Cont_i}, \mathrm{C_i})$.
          If $|\mathrm{Cint}| < |\mathrm{C_i}|$ then
                  The concept-sense $\mathrm{C_i}$ is not within the context $\mathrm{Cont_i}$ .
          EndIf
          If $|\mathrm{Cint}| = |\mathrm{C_i}|$ then
                  The concept-sense $\mathrm{C_i}$ is within the context $\mathrm{Cont_i}$ .
                  Add $\mathrm{C_i}$ to the set of possible senses associated with the document.
          EndIf
EndFor
```

Figure 3. The algorithm of Words Mapping into Concepts

1) Weight of Concept Computation

The concepts in documents are identified as a set of terms that have identified or synonym relationships, i.e., synsets in the WordNet ontology. Then, the concept frequencies $\mathrm{Cf_c}$ are calculated based on term frequency $\mathrm{tf_{tm}}$ as follows [54]:

$$\mathrm{Cf_c} = \sum_{\mathrm{t_m} \in \mathrm{r(c)}} \mathrm{tf_{tm}} \qquad (3)$$

Where r(c) is the set of different terms that belongs to concept C .

D. Text Summarization

Text summarization [25] aims at compressing an original document into a shortened version by extracting the most important information out of the document.

Extractive summary [26, 27] is used in the proposed system by extracting key text segments from the text, based on statistical analysis of individual or mixed surface level features such as word/phrase frequency, location or cue words to locate the sentences to be extracted. In the proposed system, the text summarization is performed by extracting the most relevant sentences, that are key sentences, from original document by calculating the weight of sentences [55] and then select the heigher weight. Semantic simalirity using WordNet is used to filter and refine the selected senetence to extract semantic dissimilar sentences. Figure 5 shows the algorithm of text summarization.

```
Input: document D.
Output: Set ($\mathrm{Set_{Sum}}$) of Summary Sentences ($\mathrm{Sum_S}$) in document D.
Procedure:
Split and get set of sentences (Sen) in document D
// Step1: Calculate the weight $\mathrm{Sen_w}$ of sentence (Sen)
Do While $i <= C_{Sen}$ // Count of $Sen$ in D
        For Each term T In Sen
            Get Term Weight ($\mathrm{T_w}$) as found in equation 2.
            Insert $\mathrm{T_w}$ to the set of term weight $\mathrm{Set_{tw}}$
        EndFor
        Get Length of Sentence (Len(Sen))
        Calculate $\mathrm{Sen_w} = \frac{\sum \mathrm{T_w}}{\mathrm{Len(Sen)}}$
        If $Sen$ in title or subtitle Then // if the sentence is found in distinguished location
            Calculate weight of the sentence location $\mathrm{Loc_w}$ //where $1 \le \mathrm{Loc_w} \le 1.6$
        Else
            $\mathrm{Loc_w} = 1$
        EndIf
        If $Sen$ contains special phrases Then // such as "this paper propose; this article introduce;.."
            Calculate weight of the sentence $\mathrm{Sp_w}$
        Else
            $\mathrm{Sp_w} = 1$
        EndIf
        Calculate $\mathrm{Sen_w} = \frac{\sum \mathrm{T_w}}{\mathrm{Len(Sen)}} \times \mathrm{Loc_w} \times \mathrm{Sp_w}$ // The weight of sentence
        Insert $Sen$ and its weight $\mathrm{Sen_w}$ to list $\mathrm{list_{sen}}$
        Rank the list by the $\mathrm{Sen_w}$
EndDo
// Step2: Filtering and refining the output sentence $Sen$ using semantic similarity using WordNet
Do While $j <= C_{\mathrm{list}}$ // $C_{\mathrm{list}}$ is Count of $Sen$ in $\mathrm{list_{sen}}$
            Get Semantic Similarity ($\mathrm{SemSim_{Sen(j),(j+1)}}$) of $Sen\ (j)\ and\ Sen\ (j+1)$
            Get Sentences ($\mathrm{Sum_S}$) that are Semantic Dissimilar
            Insert ($\mathrm{Sum_S}$) $in$ ($\mathrm{Set_{Sum}}$)
EndWhile
```

Figure 4. The algorithm of Text Summarization

E. User Model Acquiring

This step aims at building the user model using user behavior in the system. There are roughly two kinds of automatic way to capture a user's interest implicitly: behavior-based and history-based. Browsing histories capture the relationship between user's interests and his click history in which sufficient contextual information is already hidden in the web log. User interests [56] always constitute the most important part of the user profile in adaptive IR and filtering systems that dealt with large volumes of information.

The main purpose of this step is acquiring the interested concepts of the user in the web page (document), and then gets concept frequency that reflects the importance of concept, and finally gets the weight of concepts in the selected page. The output of this step is the weight of concepts in the selected page that can be used to build user interest model.

During the user is working through proposed system, user interests often change quite, and users are reluctant to specify all adjustments and modifications of their intents and interests. Therefore, techniques that leverage implicit approaches for gathering information about users are highly desired to update the user interests that are often not fixed.

User model in the proposed system is built in ontological representation by using domain ontology. User model is built by mapping of user's interest information and the concept in domain ontology; convert the contents of the user's interest into the form of ontology concept, and using these ontology concepts to construct user interest ontology.

Figure 6 shows the algorithm of user model acquiring.

```
Input: ($B_C$) Bag of concepts ($C_i$) in represented document D that was browsed by the user during using the system as found in section IV(B); Concepts $\mathrm{C_{Ont}}$ in domain ontology DO.
Output: User Model (UM) in ontological representation.
Procedure:
// Step 1: Acquire User Interest to build UM .
Do While $i <= C_c$    // ($C_c$) is the count of Concepts$C_i$ in the bag,
        Get concept weight $W_{Ci}$ for $C_i$ by using equation 3
        Save concept $C_i$ and its weight $W_{Ci}$ as user interest and its weight in UM
EndDo
// Step 2: Build the UM as ontological representation (user ontology).
For Each $\mathrm{C_i}$ In UM
        If $\mathrm{C_i}$ is similar to concept $\mathrm{C_{Ont}}$ in DO then
                Get Concept relations $Rel_c$ for $\mathrm{C_i}$ from DO
                Get $W_{Ci}$ for $C_i$ from UM
                Insert $\mathrm{C_{Ont}}$ and its $W_{Ci}$ to user ontology node.
                Insert $Rel_c$ of $\mathrm{C_{Ont}}$ to all related concepts
        Else
                Insert $\mathrm{C_i}$ and its $W_{Ci}$ to user ontology node.
        EndIf
EndFor
```

Figure 5. The algorithm of User Model acquiring

F. Information Retrieving

The document retrieval [57] is based on semantic similarity of the query term vector and document vector using equation 5.

$$\mathrm{sim(q,d)} = \frac{\sum_i \sum_j \mathrm{q_j\ w_i\ sim(i,j)}}{\sum_i \sum_j \mathrm{q_j w_i}} \qquad (4)$$

where w_i is term weight of concept i in the documents vector, q_j is the term weight of term j in the query vector, and $\mathrm{sim}(i, j)$ is semantic similarity [58] of the term i and term j.

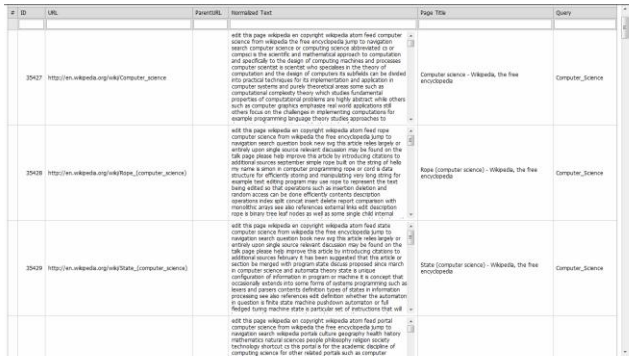
Finally, the system should arrange the retrieved documents by using the semantic similarity score of the query term vector and document vector. After building the user model; that is based on the user interest, the system uses the user model to rerank the retrieved documents. Reranking the retrieved documents user model makes the documents appears in the order as the user interest is matched. Matching between the user interest term and the documents terms is based on semantic similarity to determine the documents that the user is interested to be ranked first by semantic similarity score.

V. IMPLEMENTATION AND EXPERIMENTATION

In this section, the results of the experiments carried out to evaluate the performance of proposed system will be discussed from a quantitative point of view by running some experiments to evaluate the precision of the results. A test set collection is used to evaluate the proposed system. The test collection is a set of documents, queries and a list of relevance documents. These are used to compare the results of proposed system using the ranking strategies described.

The proposed system is implemented in ASP.Net as Web-based system using Visual Studio 2010, .NET Framework 4, and SQL Server 2008. The number of stored documents is about 3000 documents. These Web documents are about computer science domain.

Figure 7 shows the samples of the extracted texts from the collected documents.

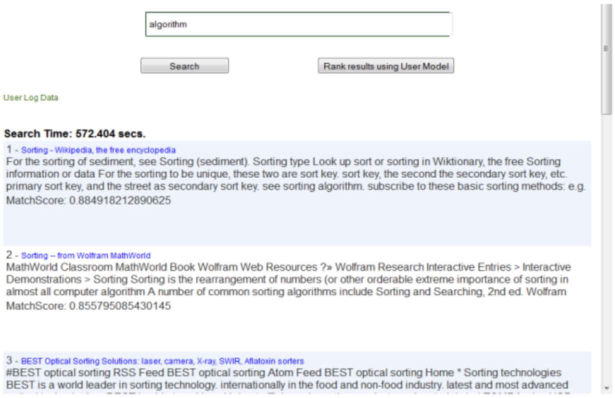


#	ID	URL	ParentURL	Normalized Text	Page Title	Query
	35427	http://en.wikipedia.org/wiki/Computer_science		edit this page wikipedia en copyright wikipedia atom feed computer science from wikipedia the free encyclopedia jump to navigation search computer science or computing science abbreviated cs or compsci is the scientific and mathematical approach to computation and specifically to the design of computing machines and processes computer scientist is scientist who specialises in the theory of computation and the design of computers its subfields can be divided into practical techniques for its implementation and application in computer systems and purely theoretical areas some such as computational complexity theory which studies fundamental properties of computational problems are highly abstract while others such as computer graphics emphasize real world applications still others focus on the challenges in implementing computations for example programming language theory studies approaches to	Computer science - Wikipedia, the free encyclopedia	Computer_Science
	35428	http://en.wikipedia.org/wiki/Rope_(computer_science)		edit this page wikipedia en copyright wikipedia atom feed rope computer science from wikipedia the free encyclopedia jump to navigation search question book new svg this article relies largely or entirely upon single source relevant discussion may be found on the talk page please help improve this article by introducing citations to additional sources september simple rope built on the string of hello my name is simon in computer programming rope or cord is data structure for efficiently storing and manipulating very long string for example text editing program may use rope to represent the text being edited so that operations such as insertion deletion and random access can be done efficiently contents description operations index split concat insert delete report comparison with monolithic arrays see also references external links edit description rope is binary tree leaf nodes as well as some single child internal	Rope (computer science) - Wikipedia, the free encyclopedia	Computer_Science
	35429	http://en.wikipedia.org/wiki/State_(computer_science)		edit this page wikipedia en copyright wikipedia atom feed state computer science from wikipedia the free encyclopedia jump to navigation search question book new svg this article relies largely or entirely upon single source relevant discussion may be found on the talk page please help improve this article by introducing citations to additional sources february it has been suggested that this article or section be merged with program state discuss proposed since march in computer science and automata theory state is unique configuration of information in program or machine it is concept that occasionally extends into some forms of systems programming such as lexers and parsers contents definition types of states in information processing see also references edit definition whether the automaton in question is finite state machine pushdown automaton or full fledged turing machine state is particular set of instructions that will	State (computer science) - Wikipedia, the free encyclopedia	Computer_Science
				edit this page wikipedia en copyright wikipedia atom feed portal computer science from wikipedia the free encyclopedia jump to navigation search wikipedia portals culture geography health history mathematics natural sciences people philosophy religion society technology shortcut cs this portal is for the academic discipline of computing science for other related portals such as computer		

Figure 7. Samples of the extracted texts from the collected documents

Figure 8 shows the results of the retrieving subsystem. This subsystem retrieves the documents based on semantic similarity between the query and the collected documents.

The improvement is measured by performing three experiments. In the experiments, relevance based evaluation method [59] is used. It uses the metrics; precision, recall, f-measure, average precession (AP) and mean average precision (MAP), to measure the performance of proposed system. In this method, proposed system is judged according to the search results' ability to satisfy an easily pleased user or hard to please user.



algorithm

Search Time: 572.404 secs.

1 - Sorting - Wikipedia, the free encyclopedia
For the sorting of sediment, see Sorting (sediment). Sorting type Look up sort or sorting in Wiktionary, the free Sorting information or data For the sorting to be unique, these two are sort key. sort key, the second the secondary sort key, etc. primary sort key, and the street as secondary sort key. see sorting algorithm. subscribe to these basic sorting methods: e.g.
MatchScore: 0.884918212890625

2 - Sorting -- from Wolfram MathWorld
MathWorld Classroom MathWorld Book Wolfram Web Resources ?» Wolfram Research Interactive Entries > Interactive Demonstrations > Sorting Sorting is the rearrangement of numbers (or other orderable extreme importance of sorting in almost all computer algorithm A number of common sorting algorithms include Sorting and Searching, 2nd ed. Wolfram
MatchScore: 0.855795085430145

3 - BEST Optical Sorting Solutions: laser, camera, X-ray, SWIR, Aflatoxin sorters
#BEST optical sorting RSS Feed BEST optical sorting Atom Feed BEST optical sorting Home * Sorting technologies BEST is a world leader in sorting technology. internationally in the food and non-food industry. latest and most advanced

Figure 6. The results of the retrieving subsystem for "algorithm" term

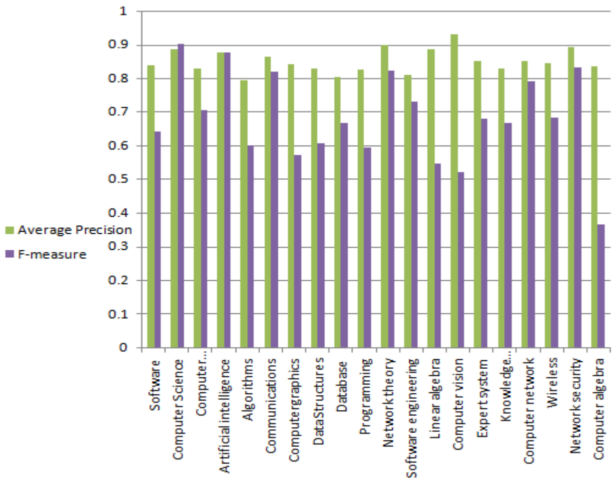


Figure 8. The average precession and f-measure for different terms

In the first experiment, figure 9 shows the average precession and f-measure that is based on different retrieving results for different user query.

The results of the different queries show that the system gives high precision during retrieving documents.

The second experiment aims at determining the importance of semantic similarity during determining the documents that are relevant to the user query. This experiment measures the performance degree when the system uses this function. It compares the recall, precision, average precision and MAP with and without using this function.

Figure 10 shows the charts of the MAP of comparison for using the semantic similarity (SemSim) between documents vector and query vector or not. This experiment shows that using semantic similarity, when the system determines the documents that are matched user query, increases the accuracy of document retrieving.

The third experiment aims at determining the importance of personalization by using generated user model (UM) during using the system. The user model is used to re-rank the retrieved documents to match the user interest.

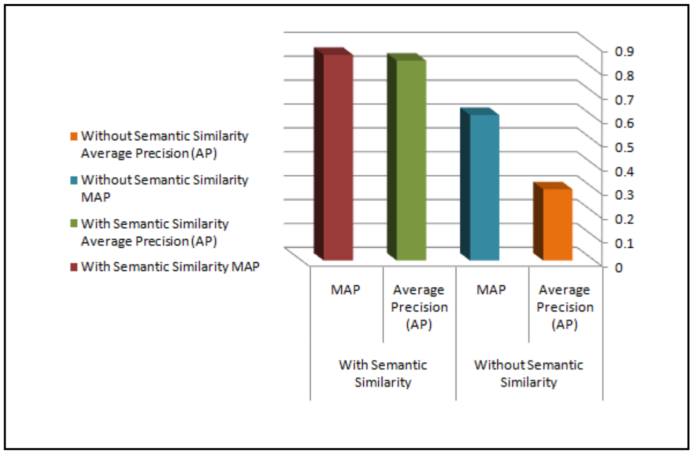


Figure 9. Coparison of using Semantic Similarity during determining the documents that are relevant to the user query

This experiment measures the performance degree when the system uses this function. It compares the recall, precision, average precision and MAP with and without using this function. Figure 11 shows the charts of the MAP of comparison for using the UM to re-rank the retrieved documents. This experiment emphasizes that using UM to realize the personalization aims at improving documents retrieving results by re-ranking the retrieved results based on user interests.

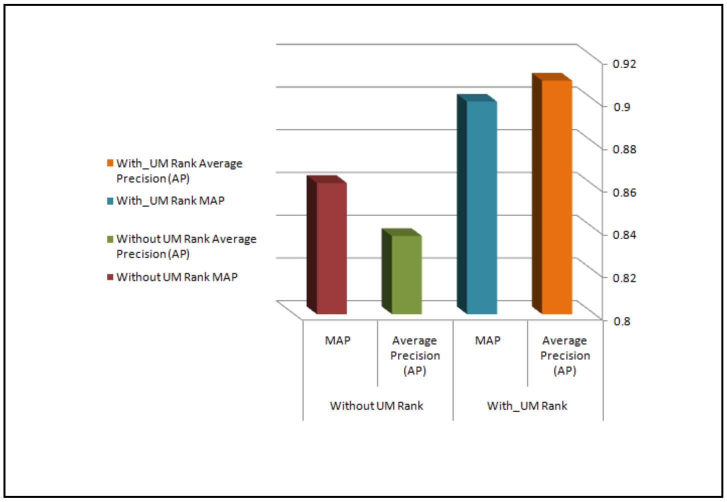


Figure 10. Coparison of using User Model during re-ranking the retrieved documents

VI. Conclusion

Aiming to solve the limitations of keyword-based models, the idea of semantic search, understood as searching by meanings rather than literal strings, has been the focus of a wide body of research in the IR and the Semantic Web communities.

A system for personalized semantic IR and summarization has been presented. The Semantic Web is a new approach for organizing information and it represents a great interest area for the international research community, but it is still far from a large-scale implementation. In this work, we have proposed and implemented the system for IR based on Semantic Web, defining a strategy for scoring and ranking results by means of a novel metric to measure semantic relatedness between terms. In the proposed system, user model, which is user interests, is used to realize the personalization. It is acquired by using concept vector model and WordNet ontology to be represented in semantic representation. In the proposed approach, summarization is based on extractive summary. Summarization is implemented by extracting the most relevant sentences, that are key sentences, from original document by calculating the weight of sentences and then select the heigher weight. Semantic simalirity using WordNet is used to filter and refine the selected .

In the system evaluation, three experiments; that are based on relevance evaluation method, show that the system can improve the accuracy of the IR because it depends on the Semantic Web tecknolgy. The system performs the summarization to allow users to find relevant full-text sources more quickly.

References

[1] Ramachandra, M. (2010). Information Retrieval. In: Web-Based Supply Chain Management and Digital Signal Processing: Methods for Effective Information Administration and Transmission. PP: 182-194. DOI: 10.4018/978-1-60566-888-8.ch014. IGI Global.

[2] Yue, X., Di, G. Yu, Y. Wang, W. & Shi, H. (2012). Analysis of the Combination of Natural Language Processing and Search Engine Technology. 2012 International Workshop on Information and Electronics Engineering (IWIEE). Procedia Engineering 29 (2012) 1636 – 1639. Elsevier Ltd.

[3] Liddy, M. (2006). Document Retrieval, Automatic. In: Encyclopedia of Language & Linguistics (Second Edition) 2006, Pages 748–755. Elsevier Ltd.

[4] MITRA, M. & CHAUDHURI, B. (2000). Information Retrieval from Documents: A Survey. Information Retrieval 2, 141–163 (2000). Kluwer Academic Publishers.

[5] Bai, J. & Nie, J. (2008). Adapting information retrieval to query contexts. Information Processing and Management 44 (2008) 1901–1922. Elsevier Ltd.

[6] Chow, T., Zhang, H. & Rahman, M.. (2009). A new document representation using term frequency and vectorized graph connectionists with application to document retrieval. Expert Systems with Applications 36 (2009) 12023–12035. Elsevier Ltd.

[7] Chen, M., Chu, H. & Chen, Y. (2010). Developing a semantic-enable information retrieval mechanism. Expert Systems with Applications 37 (2010) 322–340. Elsevier Ltd.

[8] Silva, F., Girardi, R. & Drumond, L. (2009). An Information Retrieval Model for the Semantic Web. Sixth International Conference on Information Technology: New Generations. 978-0-7695-3596-8/09, IEEE.

[9] Vester, K. & Martiny M. (2005). Information retrieval in document spaces using clustering. IMM-Thesis . Informatics and Mathematical Modelling, Technical University of Denmark, DTU.

[10] Liu. Y. (2009). On Document Representation and Term Weights in Text Classification. In: Handbook of Research on Text and Web Mining Technologies. DOI: 10.4018/978-1-59904-990-8.ch001, 1-22. IGI Global.

[11] Grossman, D. A. & Frieder, O. (2004). Information Retrieval: Algorithms And Heuristics. The Springer International Series in Engineering and Computer Science. Springer.

[12] Berners-Lee, T. (1998). Semantic Web Roadmap, W3C Semantic Web Vision Statement. http://www.w3.org/DesignIssues/Semantic.html, Last accessed on 12/09/2011.

[13] Hendler, J. (2001). Agents and the Semantic Web. IEEE Intelligent Systems, 16 (2). 30-37.

[14] Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web, Scientific American 284(5):35-43, http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html.

[15] Devedžić, V. (2005). Introduction to the Semantic Web. In: Integrated Series in Information Systems, Volume 12, 29-69, DOI: 10.1007/978-0-387-35417-0_2 . Springer-Verlag Berlin Heidelberg.

[16] Fellbaum, C. (2010). WordNet. Theory and Applications of Ontology: Computer Applications, 231, PP: 231-243, Springer Science+Business Media B.V.

[17] Maria, I. & Loke, S. (2010). The Impact of Ontology on the Performance of Information Retrieval: A Case of WordNet, In G. I. Alkhatib, D. C. Rine, Web Engineering Advancements and Trends: Building New Dimensions of Information Technology, DOI: 10.4018/978-1-60566-719-5.ch002, 24-37.

[18] Pereira, d. C., Tettamanzi, C. (2006). A.G.B.: An ontology-based method for user model acquisition. In: Ma, Z. (ed.) Soft computing in ontologies and semantic Web. Studies in fuzziness and soft computing, pp. 211–227. Springer, Heidelberg.

[19] Amine, A., Elberrichi, Z. & Simonet, M. (2010). Evaluation of Text Clustering Methods Using WordNet, The International Arab Journal of Information Technology, (7) 4.

[20] Voorhees, E. (1994). Query Expansion Using Lexical-Semantic Relations. The 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, (Dublin Ireland, 1994), 61. ACM.

[21] Anke, J. & Sundaram, D. (2009). Personalization Techniques and Their Application. In: Ang , S. & Zaphiris, P. Human Computer Interaction: Concepts, Methodologies, Tools, and Applications. DOI: 10.4018/978-1-87828-991-9.ch013. 168-176. IGI Global.

[22] Nidelkou, E., Papastathis, V., Papadogiorgaki, M., Kompatsiaris, I., Bratu, B., Ribiere, M. & Waddington, S. (2009). User Profile Modeling and Learning. In Encyclopedia of Information Science and Technology, Second Edition. DOI: 10.4018/978-1-60566-026-4.ch627. 3934-3939. IGI Global.

[23] Baishuang, Q., & Wei, Z. (2009). Student Model in Adaptive Learning System based on Semantic Web, In First International Workshop on Education Technology and Computer Science, 978-0-7695-3557-9/09, IEEE, DOI 10.1109/ETCS.2009.466.

[24] Li, F., Li, Y., Wu, Y., Zhou, K., Li, Z., Wang, X. (2008). Discovery of a User Interests on the Internet, 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, DOI 10.1109/WIIAT.2008.18, 978-0-7695-3496-1/08, IEEE.

[25] Reeve, L., Han, H. & Brooks, A. (2007). The use of domain-specific concepts in biomedical text summarization. Information Processing and Management 43 (2007) 1765–1776. Elsevier Ltd.

[26] Gupta, V. & Lehal, G. (2010). A Survey of Text Summarization Extractive Techniques. JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE, VOL. 2, NO. 3, AUGUST 2010. ACADEMY PUBLISHER.

[27] Kyoomarsi, F., Khosravi, H. Eslami, E., Dehkordy, P. & Tajoddin, A. (2008). Optimizing Text Summarization Based on Fuzzy Logic. Seventh IEEE/ACIS International Conference on Computer and Information Science. 978-0-7695-3131-1/08, IEEE.

[28] Gupta, V. & Lehal, G. (2009). A Survey of Text Mining Techniques and Applications. JOURNAL OF EMERGING TECHNOLOGIES IN WEB INTELLIGENCE, VOL. 1, NO. 1, AUGUST, ACADEMY PUBLISHER.

[29] Gauch, S., Speretta, M. & Pretschner, M. (2007), Ontology-Based User Profiles for Personalized Search, DOI10.1007/978-0-387-37022-4, Springer US.

[30] Tanudjaja, F. & Mui, L. (2002) Persona: A Contextualized and Personalized Web Search. Proc 35 th Hawaii Intl. Conf. on System Sciences.

[31] The Open Directory Project (ODP). http://dmoz.org.

[32] Pitkow, J., Schütze, H. & Cass, T.. (2002), Personalized search. CACM 2002; 45(9):50-55.

[33] McKeown, K., Elhadad, N. & Hatzivassiloglou, V..(2003), Leveraging a common representation for personalized search and summarization in a medical digital library. In Proceedings of the 3 rd ACM/IEEE-CS joint conference on Digital libraries 2003; 159-170.

[34] Liu, F., Yu, C. & Meng, W. (2002) Personalized web search by mapping user queries to categories. In Proceedings CIKM'02 2002; 558-565.

[35] Sugiyama, K., Hatano, K. & Yoshikawa, M. (2004), Adaptive web search based on user profile constructed without any effort from users. In Proceedings 13 th Intl. Conf. on World Wide Web 2004; 675-684.

[36] Xiangwei, M., Yan, C. & Nan, L. (2009), Modeling of Personalized Recommendation System Based on Ontology, 978-1-4244-4639-1/09, IEEE.

[37] Trong, D., Mohammed, N, Delong, L., & Geun, J. (2009), A Collaborative Ontology-Based User Profiles System, N.T. Nguyen, R. Kowalczyk, and S.-M. Chen (Eds.): ICCCI 2009, LNAI 5796, pp. 540–552, Springer-Verlag Berlin Heidelberg.

[38] Jina, H. , Ninga, X., Jiab, W., Wua, H. & Luc, G.. (2008), Combining weights with fuzziness for intelligent semantic web search, Knowledge-Based Systems 21 (2008) 655–665, 0950-7051, Elsevier.

[39] Yang. Q., Sun, J., Li, Y. & Ca, K. (2010), Domain Ontology-based personalized recommendation research, 978-1-4244-5824-0, IEEE.

[40] Pereira, C. & Tettamanzi, A. (2006), An Evolutionary Approach to Ontology-Based User Model Acquisition, V. Di Ges´u, F. Masulli, and A. Petrosino (Eds.): WILF 2003, LNAI 2955, pp. 25–32, c_Springer-Verlag Berlin Heidelberg.

[41] Reformat, M. Koosha, S. (2009). Updating User Profile using Ontology-based Semantic Similarity, FUZZ_IEEE 2009, Korea, August 20-24, 978-1-4244-3597-5, IEEE.

[42] Dıaz, A. & Gervas, P. (2007). User-model based personalized summarization. Information Processing and Management 43 (2007) 1715–1734. Elsevier Ltd.

[43] Lv, G., Zheng, C. & Zhang, L. (2009). Text Information Retrieval Based on Concept Semantic Similarity. 2009 Fifth International Conference on Semantics, Knowledge and Grid. 978-0-7695-3810-5/09. IEEE.

[44] Rinaldi, A. M. (2009). An ontology-driven approach for semantic information retrieval on the Web. ACM Trans. Internet Technol, 9, 3, Article 10 (July 2009), 24 pages. DOI = 10.1145/1552291.1552293 http://doi.acm.org/10.1145/1552291.1552293

[45] Huang, G. & Zhang, X. (2010). Text Retrieval based on Semantic Relationship. 978-1-4244-7161-4/10, IEEE.

[46] Gauch, S., Speretta, M. & Pretschner, A. (2007). ONTOLOGY-BASED USER PROFILES FOR PERSONALIZED SEARCH. In: Ontologies A Handbook of Principles, Concepts and Applications in Information Systems, Volume 14, 2007, DOI: 10.1007/978-0-387-37022-4. Springerlink.

[47] Majestic-12: Projects : C# HTML parser (.NET). http://www.majestic12.co.uk/projects/html_parser.php.

[48] wordnetdotnet - Revision 262. http://wordnetdotnet.googlecode.com/svn/trunk/Projects/Thanh/.

[49] Bai, R., Wang, X. & Liao, J. (2010). Extract Semantic Information from WordNet to Improve Text Classification Performance. AST/UCMA/ISA/ACN 2010, LNCS 6059, pp. 409–420. Springer-Verlag Berlin Heidelberg.

[50] Gharib, T., Fouad, M. & Aref, M. (2010). Fuzzy Document Clustering Approach using WordNet Lexical Categories. In: Advanced Techniques in Computing Sciences and Software Engineering. DOI 10.1007/978-90-481-3660-5, Springer Science+Business Media.

[51] Tarek, G., Fouad, M. & Aref, M. (2008). Web Document Clustering Approach using WordNet Lexical Categories and Fuzzy Clustering. Proceedings of International Workshop on Data Mining and Artificial Intelligence (DMAI' 08), 24 December, 2008, Khulna, Bangladesh. 1-4244-2136-7/08, IEEE.

[52] Jones, K. (2004). A Statistical Interpretation of Term Specificity and its Application to Retrieval. Journal of Documentation, 60 (5), p.493-502.

[53] B. Fatiha, B. Mohand, T. Lynda, D. Mariam. (2010). Using WordNet for Concept-Based Document Indexing in Information Retrieval, SEMAPRO: The Fourth International Conference on Advances in Semantic Processing, Pages: 151 to 157, IARIA.

[54] Dragoni, M., Pereira, C. & Tettamanzi, A. (2010). An Ontological Representation of Documents and Queries for Information Retrieval Systems, IEA/AIE 2010, Part II, LNAI 6097, pp. 555–564, Springer-Verlag Berlin Heidelberg.

[55] Xu, X. (2009). Research on Automatic Summarization System based on topic partition. 2009 International Conference on Web Information Systems and Mining. 978-0-7695-3817-4/09, IEEE.

[56] Nidelkou, E., Papastathis, V., Papadogiorgaki, M., Kompatsiaris, I., Bratu, B., Ribiere, M. & Waddington, S. (2009). User Profile Modeling and Learning. In Encyclopedia of Information Science and Technology, Second Edition. DOI: 10.4018/978-1-60566-026-4.ch627. 3934-3939. IGI Global.

[57] Lv, G., Zheng. C. & Zhang, L. (2009). Text Information Retrieval Based on Concept Semantic Similarity. Fifth International Conference on Semantics, Knowledge and Grid. 978-0-7695-3810-5/09, IEEE.

[58] Saruladha, K., Aghila.G. & Raj, S. (2010). A Survey of Semantic Similarity Methods for Ontology based Information Retrieval. Second International Conference on Machine Learning and Computing. 978-0-7695-3977-5/10, IEEE.

[59] Ali, R. & Beg, M. (2011). An overview of Web search evaluation methods. Computers and Electrical Engineering 37 (2011) 835–848. Elsevier Ltd.

AUTHORS PROFILE

Salah T. Babekr is an associate professor of Computer Engineering in College of Computers and Information Technology, Taif University. He is PhD holder for 16 years as a Computer Engineer bilingual Russian and English with extensive experience in administration and project control,analysis, design, consultancy, development and implementation of software, organization, establishment and improvement ofInternet band networks security, quality control on software products, implementation of best development practices, teamwork and support.

Khaled M. Fouad has received his PhD and Master degree of AI, and expert systems in of computer engineering from the faculty of engineering AlAzhar University in Egypt. He is working now as assistant professor in Taif University in Kingdom of Saudi Arabia (KSA) and is researcher in Central Laboratory of Agriculture Expert Systems (CLAES) in Egypt. His current research interests focus on semantic web, text mining, clustering and expert systems.

Naveed Arshad has completed his Ph.D. from University of Colorado at Boulder, USA. Before joining LUMS, Dr Naveed Arshad has worked with ABN AMRO Global IT Systems, Pakistan International Airline. He is part of the Software Engineering Research Group (SERG) at LUMS. This group is undertaking research in various areas of software engineering such as engineering of autonomic systems, conceptual modeling, large scale systems development, etc.

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

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 block cipher by extending the analysis of a Novel Block Cipher Involving a Key bunch Matrix and a Key-based Permutation and Substitution. Here we have include and additional key matrix, which is supplemented with xor operation. The cryptanalysis carried out in this investigation clearly indicates that this cipher cannot be broken by any attack.**

***Keywords-** Key; key bunch matrix; encryption; decryption; permutation; substitution; avalanche effect; cryptanalysis; xor operation*

I. Introduction

In a recent investigation [1], we have developed a block cipher involving a key bunch matrix and including a pair of functions, called Permute() and Substitute(). In this analysis, we have seen that the permutation and the substitution, which depend effectively on a key, strengthen the cipher in a remarkable manner. This is all on account of the fact that the permutation and the substitution, induced into the plaintext at each and every stage in the iteration process, causes confusion and diffusion.

In the present investigation, our objective is to modify the afore-mentioned block cipher by introducing an additional key matrix supplemented with xor operation. The basic equation governing the encryption of this cipher can be written in the form

$$C = [\, c_{ij} \,] = ([\, e_{ij} \times p_{ij} \,] \bmod 256) \oplus F, \quad i=1 \text{ to } n,\ j = 1 \text{ to } n. \qquad (1.1)$$

The corresponding equation describing decryption can be written in the form

$$P = [\, p_{ij} \,] = [\, d_{ij} \times (C \oplus F)_{ij} \,] \bmod 256, \quad i=1 \text{ to } n,\ j = 1 \text{ to } n. \qquad (1.2)$$

Here, our interest is to examine, how the additional key matrix, F, would strengthen the cipher when supported by permuted and substitution.

Let us now present the plan of the paper. In section 2, we introduce the development of the cipher. Here, we depict the flowcharts and write the algorithms required in this investigation. Then, we mention the basic ideas of the key based permutation and substitution. In section 3, we mention an illustration of the cipher, and discuss the avalanche effect. We study the cryptanalysis, in section 4. Finally, we deal with the computations carried out in this analysis, and draw conclusions, in section 5.

II. Development of the Cipher

Consider a plaintext, which can be written in the form of a square matrix P, given by

$$P = [\, p_{ij} \,], \ i=1 \text{ to } n,\ j=1 \text{ to } n. \qquad (2.1)$$

Let us take a key bunch matrix E, given by

$$E = [\, e_{ij} \,], \ i=1 \text{ to } n,\ j=1 \text{ to } n. \qquad (2.2)$$

On using the concept of the multiplicative inverse [2], we get d_{ij} corresponding to each e_{ij} . Thus we have the decryption key bunch matrix D, given by

$$D= [\, d_{ij} \,], \ i=1 \text{ to } n,\ j=1 \text{ to } n. \qquad (2.3)$$

Here, it is to be noted that, all the e_{ij} and d_{ij} are odd numbers which lie in the interval [1-255].

The flowcharts concerned to the encryption and the decryption are drawn in Figs. 1 and 2.

The corresponding algorithms for the encryption and the decryption are 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.   $p_{ij} = (e_{ij} \times p_{ij})$ mod 256
     }
     }
6.   P=[ $p_{ij}$ ]$\oplus$ F
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.   $c_{ij} = [d_{ij} \times (c_{ij} \oplus f_{ij})]$ mod 256
     }
     }
9.   C=[ $c_{ij}$ ]
     }
10.  P=C
11.  Write (P)
```

In this analysis, r denotes the number of rounds in the iteration process, and it is taken as 16.

The functions Permute() and Substitute(), which are utilized in encryption, depend upon a key. Let us choose the key, K, in the form

$$K = \begin{bmatrix} 156 & 14 & 33 & 96 \\ 253 & 107 & 110 & 127 \\ 164 & 10 & 5 & 123 \\ 174 & 202 & 150 & 94 \end{bmatrix}$$

Keeping the serial numbers and the order of the elements in the key, in view, we construct a table of the form given in Table-1.

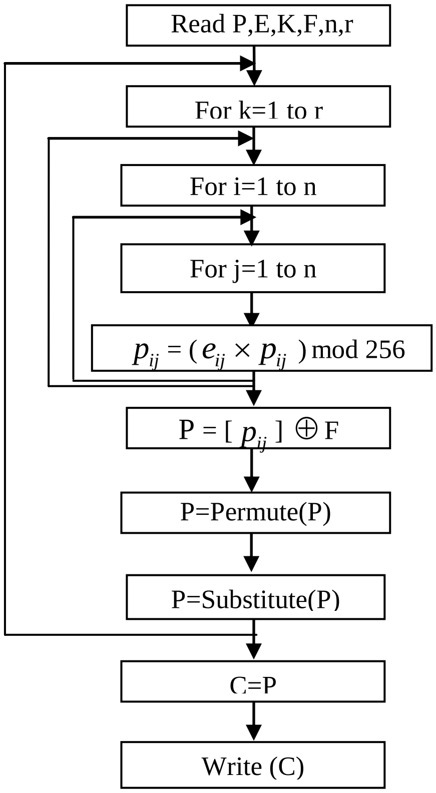


Fig.1 Flowchart for Encryption

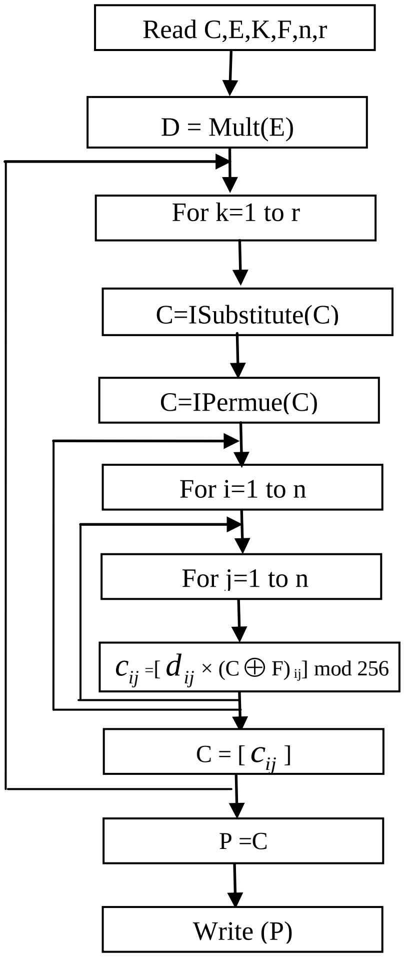


Fig.2 Flowchart for Decryption

Table-1. 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

The process of permutation can be explained as follows.

Let x_i, i=1 to 16, be a set of 16 numbers. As the table is suggesting (looking at the first row and the third row), we interchange x_1 with x_{12}, x_2 with x_3, x_4 with x_6, x_5 with x_{16}, x_7 with x_8, x_9 with x_{13}, and x_{14} with x_{15}. It may be noted here that we need not interchange any other numbers as they are already subjected to change in a way. Keeping this basic idea in view, let us now consider the plaintext matrix P = [p_{ij}], i=1 to n, j=1 to n, (after xoring with F) in any round of the iteration process. Considering the first two rows of this matrix and representing the elements p_{ij} in their binary form, and writing the binary bits in the vertical manner, we get a matrix of size 16xn. On dividing this matrix into sub-matrices, where each one of size 16x16, and performing the interchange of rows (firstly) and columns (subsequently), as is done in the case of numbers x_i , i=1 to 16, we get the corresponding permuted matrix, in the case of each sub-matrix. On applying the same procedure for the other sub-matrices also, we ultimately get n/16 sub-matrices. On representing the binary bits in terms of decimal numbers (converting 8 binary bits in a row as a decimal number), we get a 2xn matrix. On adopting the same procedure on the subsequent pairs of this matrix, we complete the permutation process. However, it is to be remembered that n must be divisible by 16. In case, if n<16, that is say, n=4, then a plaintext matrix of size 4x4 can be written as a matrix of size 8x16, by writing each decimal number as binary bits in a column. Then the procedure of swapping, applied for numbers, can be applied here, for rows firstly and for columns nextly.

However, in the case of rows, we restrict our interchanging process only to 8 rows. Then, on representing the binary bits in terms of decimal numbers (considering the bits in a row-wise manner) we get the permuted matrix. This completes t he process of permutation.

The process of substitution can be mentioned as follows. In the EBCDIC code, the characters can be represented in terms of a table of size 16x16, containing numbers 0 to 255, in a sequential manner. On swapping rows, firstly, and columns, nextly, as it is already done in the case of the numbers x_1 to x_{16}, we get a new table (see Table-2).

On using the Table-2, we perform substitution, by noting the correspondence between the number in the plaintext, the number in the EBCDIC table and hence the number in the substitution table. For clarity if this substitution process, we refer to [1].

The functions IPermute() and ISubstitute(), used in the decryption process, denote the reverse processes of the Permute() and the Substitute(). The function Mult() is used to find the decryption key bunch matrix D for the given E.

III. Illustration of the Cipher and the Avalanche Effect

Consider the plaintext given below.

Dear Madam! I have received your letter. Please do not run away from our country in that manner. I am coming within this month. I will not continue my Ph.D. programme. I may leave this research activity but I cannot leave you. It is indeed a surprise. Though there was no response from the selection committee for a span of one year, very recently I got selected in our country for IAS. I think I am lucky. Tell you father and mother about this news and tell them in a nice manner that you are running p8third month. I hope that all these issues will end up very soon and we will become one undoubtedly. Tell my father and mother that I am coming there. Yours loving husband

(3.1)

Let us focus our attention on the first 16 characters of this plaintext. Thus we have

Dear Madam! I ha **(**3.2)

On using the EBCDIC code, we get

$$P = \begin{bmatrix} 196 & 133 & 129 & 153 \\ 64 & 212 & 129 & 132 \\ 129 & 148 & 79 & 64 \\ 201 & 64 & 136 & 129 \end{bmatrix}. \qquad (3.3)$$

Let us take the key bunch matrix E in the form

$$E = \begin{bmatrix} 199 & 23 & 67 & 211 \\ 67 & 91 & 93 & 5 \\ 11 & 19 & 51 & 145 \\ 109 & 223 & 251 & 5 \end{bmatrix}. \qquad (3.4)$$

On using the concept of the multiplicative inverse, we have the decryption key bunch matrix D in the form

$$D = \begin{bmatrix} 247 & 167 & 107 & 91 \\ 107 & 211 & 245 & 205 \\ 163 & 27 & 251 & 113 \\ 101 & 31 & 51 & 205 \end{bmatrix}. \qquad (3.5)$$

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
51	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

TABLE-2 KEY BASED SUBSTITUTION

The additional key matrix F is taken in the form

$$F = \begin{bmatrix} 222 & 243 & 122 & 45 \\ 56 & 22 & 100 & 99 \\ 104 & 76 & 45 & 11 \\ 9 & 22 & 25 & 67 \end{bmatrix}. \qquad (3.6)$$

Now, on making use of the plaintext P, the encryption key bunch matrix E and the additional key matrix F, and applying the encryption algorithm, given in section 2, we get the ciphertext matrix C in the form

$$C = \begin{bmatrix} 88 & 2 & 165 & 241 \\ 47 & 226 & 95 & 110 \\ 214 & 121 & 129 & 163 \\ 104 & 97 & 195 & 215 \end{bmatrix}. \qquad (3.7)$$

On using this C, the F, and the decryption key bunch matrix D, given by (3.5), and the decryption algorithm, given in section 2, we get back the plaintext P, which is in the form (3.3).

Now let us study the avalanche effect. On replacing the 4th row 4th column element, 129 in the plaintext (3.3) by 193, we have a one binary bit change in the plaintext. On using this modified plaintext, the E, the F, and the encryption algorithm, we get the new ciphertext in the form, given by (3.8).

On comparing (3.8) and (3.7), after converting them into their binary form, we notice that there is a change of 72 bits out of 128 bits. This shows that the cipher is a strong one.

$$C = \begin{bmatrix} 1 & 198 & 243 & 34 \\ 189 & 43 & 134 & 140 \\ 89 & 195 & 102 & 168 \\ 149 & 148 & 254 & 196 \end{bmatrix}. \qquad (3.8)$$

Now, let us have one binary bit change in the key bunch matrix E. To this end, we replace the 3rd row 2nd column element 19 in E by 18. On using this modified E, the original P, given by (3.3), and the F, given by (3.6), and applying the encryption algorithm, we get the corresponding ciphertext C, in the form

$$C = \begin{bmatrix} 154 & 160 & 102 & 158 \\ 173 & 29 & 134 & 243 \\ 236 & 190 & 127 & 195 \\ 209 & 188 & 48 & 241 \end{bmatrix}. \qquad (3.9)$$

Now, let us convert (3.7) and (3.9) into their binary form and compare them. From this, we find that these two ciphertexts differ by 74 bits out of 128 bits. This also shows that the cipher is having appreciable strength.

Now, on making use of the plaintext P, the encryption key bunch matrix E and the additional key matrix F, and applying the encryption algorithm, given in section 2, we get the ciphertext matrix C in the form

On using this C, the F, and the decryption key bunch matrix D, given by (3.5), and the decryption algorithm, given in section 2, we get back the plaintext P, which is in the form (3.3).

Now let us study the avalanche effect. On replacing the 4th row 4th column element, 129 in the plaintext (3.3) by 193, we have a one binary bit change in the plaintext. On using this modified plaintext, the E, the F, and the encryption algorithm, we get the new ciphertext in the form

Now, let us convert (3.7) and (3.9) into their binary form and compare them. From this, we find that these two ciphertexts differ by 74 bits out of 128 bits. This also shows that the cipher is having appreciable strength.

IV. CRYPTANALYSIS

The study of cryptanalysis plays a prominent role in the development of every cipher. The different types of attacks that are available in the literature of cryptography are

1. Ciphertext only attack (Brute force attack),
2. Known plaintext attack,
3. Chosen plaintext attack, and
4. Chosen ciphertext attack.

Generally, every algorithm is designed [2] such that it withstands the first two attacks. The cipher is also examined in a thorough manner, by using all possible intuitive ideas, in the case of the latter two attacks.

In this analysis, we have the key bunch matrix E, whose size is nxn. Besides this, we have the additional key matrix F, whose size is also nxn. In addition to these two, we have the key matrix K which is used in the development of permutation and substitution processes. In view of all these three, the size of the key space is

$$2^{7n^2} \times 2^{8n^2} \times 2^{128} = 2^{7n^2+8n^2+128} = 2^{15n^2+128}$$

$$= \left(2^{10}\right)^{\left(1.5n^2+12.8\right)} \approx \left(10^3\right)^{\left(1.5n^2+12.8\right)} = 10^{4.5n^2+38.4}$$

On assuming that, we require 10^{-7} seconds for computation with one set of-n keys in the key space, then the time required for all such possible set s 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 our present analysis, as 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 time is very large, it is simply impossible to break this cipher by the brute force attack.

Let us now examine the known plaintext attack. In order to carry out this approach, we have plaintext and ciphertext pairs, as many as we want, at our disposal. If we focus our attention on only one round of the iteration process, that is if r=1, then the basic equations governing the encryption process are given by

P = ([e_{ij} × p_{ij}] mod 256) ⊕ F, 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)

In this attack, the ciphertext C in (4.4), is known to us. On using this one, we can know the P, occurring in the left side of (4.3). As key is unknown, we do not know ISubstitute(). Hence P occurring on the right hand side of (4.3) cannot be determined. Hence this cipher cannot be broken by the known plaintext attack. Luckily, if key K is known (a very stray case), then we can obtain P, occurring on the left hand side of (4.1). Then, though p_{ij} is known to us, we cannot determine the e_{ij} , by any means, as the equation (4.1) is containing several unknowns related to F, and is including mod and xor operations. Thus this cipher cannot be broken by the known plaintext attack, even when r=1, and the key matrix K, used in the permutation process is known to the attacker.

In view of the equations, involved in the encryption process, we do not find any possibility to choose either a plaintext or a ciphertext for breaking this cipher.

In the light of the above facts, we conclude that this cipher is a very strong one.

V. Computations and Conclusions

In this investigation, we have developed a block cipher involving a key bunch matrix and an additional key matrix, and involving key-based permutation and substitution. The strength of the cipher is highly remarkable due to permutation and substitution, and it is further supplemented with the additional key matrix.

The programs for encryption and decryption are written in Java.

In order to carry out the encryption of the entire plaintext, given by (3.1), we use a large size encryption key bunch matrix EK of size 16x16. Along with this, we have taken an additional key matrix FK, which is also of the same size 16x16. The EK and FK are given below in (5.1) and (5.2).

$$EK = \begin{bmatrix}
125 & 171 & 129 & 101 & 141 & 225 & 251 & 47 & 69 & 123 & 121 & 65 & 177 & 5 & 131 & 243 \\
213 & 29 & 227 & 127 & 61 & 107 & 195 & 145 & 83 & 89 & 221 & 167 & 151 & 79 & 125 & 167 \\
3 & 41 & 213 & 161 & 35 & 131 & 203 & 125 & 125 & 41 & 177 & 231 & 15 & 21 & 93 & 111 \\
209 & 83 & 65 & 203 & 183 & 163 & 165 & 59 & 123 & 15 & 113 & 157 & 249 & 243 & 171 & 113 \\
195 & 45 & 63 & 23 & 191 & 197 & 25 & 129 & 177 & 151 & 221 & 217 & 21 & 173 & 31 & 185 \\
103 & 17 & 47 & 3 & 223 & 223 & 167 & 13 & 43 & 241 & 173 & 117 & 31 & 113 & 227 & 93 \\
37 & 219 & 195 & 175 & 199 & 83 & 79 & 217 & 233 & 217 & 169 & 253 & 127 & 75 & 163 & 243 \\
215 & 111 & 79 & 159 & 193 & 5 & 231 & 117 & 55 & 55 & 63 & 119 & 249 & 205 & 193 & 13 \\
231 & 243 & 199 & 115 & 201 & 67 & 173 & 195 & 19 & 191 & 17 & 145 & 219 & 155 & 39 & 241 \\
251 & 223 & 231 & 95 & 105 & 201 & 119 & 51 & 181 & 229 & 181 & 167 & 247 & 153 & 225 & 149 \\
37 & 183 & 253 & 177 & 117 & 33 & 17 & 231 & 163 & 83 & 195 & 157 & 223 & 13 & 95 & 95 \\
183 & 241 & 95 & 53 & 247 & 117 & 169 & 23 & 27 & 107 & 85 & 167 & 215 & 171 & 203 & 139 \\
49 & 221 & 127 & 69 & 127 & 245 & 73 & 3 & 113 & 125 & 237 & 45 & 55 & 115 & 241 & 221 \\
213 & 85 & 21 & 15 & 21 & 205 & 85 & 203 & 105 & 235 & 155 & 5 & 105 & 153 & 109 & 135 \\
223 & 133 & 239 & 181 & 127 & 157 & 77 & 243 & 17 & 129 & 133 & 161 & 11 & 65 & 93 & 169 \\
91 & 59 & 171 & 201 & 53 & 91 & 31 & 169 & 203 & 113 & 181 & 125 & 151 & 165 & 245 & 51
\end{bmatrix} \qquad (5.1)$$

and

$$FK = \begin{bmatrix} 91 & 46 & 145 & 165 & 147 & 49 & 59 & 169 & 175 & 168 & 103 & 104 & 148 & 178 & 111 & 70 \\ 10 & 203 & 14 & 102 & 66 & 123 & 116 & 111 & 21 & 15 & 196 & 54 & 130 & 244 & 239 & 244 \\ 196 & 118 & 21 & 164 & 34 & 129 & 100 & 230 & 170 & 7 & 247 & 118 & 79 & 59 & 79 & 221 \\ 38 & 189 & 221 & 142 & 11 & 39 & 142 & 255 & 168 & 49 & 78 & 150 & 157 & 183 & 101 & 161 \\ 145 & 139 & 227 & 131 & 17 & 224 & 116 & 99 & 108 & 144 & 176 & 161 & 50 & 35 & 105 & 20 \\ 150 & 211 & 123 & 240 & 174 & 55 & 101 & 210 & 141 & 87 & 83 & 246 & 46 & 70 & 53 & 46 \\ 108 & 46 & 80 & 112 & 172 & 232 & 228 & 69 & 97 & 232 & 166 & 102 & 70 & 63 & 94 & 18 \\ 18 & 214 & 79 & 151 & 79 & 250 & 10 & 116 & 81 & 115 & 228 & 77 & 121 & 12 & 153 & 167 \\ 131 & 133 & 132 & 246 & 53 & 94 & 132 & 39 & 151 & 183 & 207 & 36 & 194 & 222 & 227 & 70 \\ 141 & 193 & 91 & 210 & 120 & 146 & 152 & 224 & 202 & 110 & 34 & 0 & 73 & 176 & 4 & 0 \\ 198 & 2 & 215 & 20 & 141 & 203 & 183 & 92 & 214 & 217 & 37 & 140 & 141 & 161 & 211 & 248 \\ 25 & 70 & 49 & 234 & 95 & 31 & 25 & 99 & 200 & 248 & 251 & 243 & 15 & 149 & 206 & 78 \\ 116 & 135 & 103 & 157 & 37 & 64 & 242 & 116 & 246 & 219 & 17 & 71 & 249 & 157 & 127 & 34 \\ 17 & 148 & 51 & 32 & 121 & 45 & 163 & 192 & 14 & 166 & 62 & 211 & 64 & 156 & 40 & 50 \\ 220 & 210 & 244 & 208 & 41 & 113 & 132 & 254 & 115 & 174 & 22 & 231 & 196 & 188 & 67 & 126 \\ 121 & 96 & 4 & 234 & 70 & 102 & 186 & 145 & 133 & 69 & 222 & 158 & 239 & 30 & 75 & 146 \end{bmatrix} \tag{5.2}$$

The entire plaintext, given by (3.1), is divided into 3 blocks, wherein each block is of size 16x16. In the 3rd block, we have appended 95 zeroes as characters, so that we make it a complete block. On using EK, FK, in the place of E and F, in the encryption algorithm, we carry out the encryption process 3 times, so that the complete plaintext is converted into the corresponding ciphertext. This ciphertext is given by (5.3).

The EK and FK are encrypted by using the E, the F, and applying the encryption algorithm. The resulting ciphertexts of the keys EK and FK are as follows, in (5.4) and (5.5).

These are transmitted to the receiver by the sender. In addition to these, the key bunch matrix E, the additional key matrix F, the key K used in the processes permutation and substitution are sent by the sender to the receiver, in a secure manner. The number of additional characters appended in the last block is also informed to the receiver.

38	148	35	233	157	94	147	31	252	129	27	155	11	231	166	169
227	99	59	188	89	55	67	212	75	208	216	147	227	18	217	166
109	69	91	40	101	193	98	1	60	135	125	188	51	73	45	59
24	125	41	75	17	36	157	134	53	226	20	204	25	248	99	200
6	170	236	237	98	229	205	189	41	162	91	140	73	182	9	78
212	221	116	182	225	53	203	180	252	188	235	233	247	88	149	86
66	213	88	95	157	43	93	182	30	177	134	138	98	231	199	79
241	156	13	75	4	115	91	86	228	10	138	18	244	149	8	0
81	145	34	125	247	160	62	115	40	239	253	1	17	37	113	74
116	83	51	136	116	57	107	126	236	164	167	104	18	200	83	18
0	126	111	20	47	152	172	104	81	138	65	210	181	145	206	78
53	44	149	99	233	58	76	209	40	33	42	224	179	221	160	90
223	201	57	138	120	212	159	105	173	176	164	233	198	118	108	126
134	55	93	84	231	232	250	150	28	201	170	51	8	112	254	139
28	37	30	175	46	13	75	10	139	102	118	118	39	97	121	241
44	104	224	190	16	209	148	231	150	83	22	252	166	156	19	203

$$\begin{matrix}
94 & 156 & 21 & 226 & 84 & 41 & 36 & 223 & 26 & 192 & 97 & 94 & 125 & 189 & 59 & 237 \\
231 & 99 & 144 & 43 & 241 & 159 & 157 & 217 & 34 & 186 & 20 & 244 & 191 & 43 & 85 & 152 \\
5 & 155 & 0 & 98 & 149 & 218 & 53 & 3 & 172 & 204 & 84 & 118 & 108 & 12 & 55 & 76 \\
151 & 99 & 60 & 175 & 243 & 251 & 17 & 35 & 228 & 141 & 243 & 153 & 49 & 174 & 240 & 181 \\
94 & 139 & 65 & 240 & 120 & 95 & 148 & 93 & 108 & 245 & 248 & 240 & 102 & 58 & 108 & 8 \\
40 & 96 & 157 & 14 & 62 & 24 & 190 & 164 & 167 & 227 & 8 & 251 & 109 & 45 & 97 & 187 \\
43 & 2 & 132 & 243 & 32 & 40 & 78 & 173 & 103 & 176 & 38 & 58 & 163 & 124 & 185 & 233 \\
242 & 73 & 171 & 62 & 202 & 177 & 208 & 202 & 3 & 213 & 221 & 121 & 115 & 38 & 96 & 154 \\
89 & 203 & 141 & 46 & 252 & 18 & 149 & 99 & 165 & 125 & 22 & 239 & 46 & 36 & 52 & 54 \\
121 & 175 & 143 & 164 & 45 & 196 & 2 & 163 & 2 & 80 & 208 & 112 & 229 & 196 & 97 & 56 \\
253 & 217 & 107 & 97 & 162 & 180 & 229 & 220 & 142 & 161 & 225 & 229 & 109 & 239 & 177 & 19 \\
98 & 92 & 205 & 89 & 31 & 151 & 4 & 68 & 116 & 9 & 171 & 217 & 100 & 53 & 247 & 110 \\
225 & 136 & 232 & 70 & 124 & 192 & 22 & 140 & 188 & 211 & 109 & 134 & 97 & 203 & 41 & 238 \\
51 & 123 & 20 & 89 & 179 & 222 & 63 & 93 & 70 & 92 & 203 & 170 & 185 & 35 & 52 & 90 \\
167 & 116 & 159 & 7 & 174 & 158 & 159 & 82 & 118 & 160 & 73 & 39 & 110 & 31 & 75 & 179 \\
26 & 247 & 244 & 28 & 163 & 166 & 11 & 144 & 203 & 86 & 180 & 33 & 200 & 138 & 201 & 181
\end{matrix}$$

$$\begin{matrix}
12 & 181 & 81 & 91 & 149 & 203 & 173 & 223 & 77 & 216 & 176 & 223 & 5 & 211 & 181 & 161 \\
39 & 123 & 8 & 93 & 233 & 6 & 230 & 76 & 127 & 189 & 226 & 144 & 34 & 83 & 134 & 221 \\
251 & 138 & 37 & 98 & 24 & 37 & 227 & 199 & 123 & 24 & 134 & 133 & 140 & 186 & 132 & 8 \\
114 & 170 & 231 & 225 & 178 & 141 & 116 & 190 & 89 & 63 & 243 & 59 & 200 & 61 & 84 & 247 \\
49 & 192 & 79 & 209 & 105 & 47 & 144 & 254 & 206 & 42 & 178 & 254 & 228 & 204 & 220 & 49 \\
9 & 167 & 2 & 213 & 179 & 134 & 249 & 23 & 193 & 100 & 28 & 205 & 62 & 69 & 119 & 91 \\
109 & 6 & 127 & 88 & 45 & 2 & 147 & 226 & 135 & 49 & 240 & 209 & 246 & 206 & 224 & 125 \\
71 & 185 & 146 & 5 & 72 & 206 & 99 & 67 & 233 & 152 & 192 & 253 & 13 & 154 & 215 & 36 \\
67 & 22 & 233 & 65 & 248 & 236 & 180 & 223 & 114 & 45 & 4 & 195 & 106 & 215 & 123 & 135 \\
0 & 31 & 236 & 22 & 169 & 81 & 206 & 62 & 34 & 170 & 194 & 54 & 77 & 233 & 160 & 141 \\
153 & 196 & 90 & 225 & 27 & 31 & 225 & 226 & 94 & 179 & 143 & 130 & 195 & 44 & 64 & 82 \\
255 & 30 & 13 & 203 & 62 & 194 & 21 & 17 & 106 & 201 & 56 & 2 & 71 & 210 & 24 & 231 \\
17 & 119 & 167 & 107 & 156 & 63 & 62 & 233 & 182 & 46 & 160 & 38 & 58 & 50 & 165 & 173 \\
78 & 175 & 80 & 75 & 113 & 233 & 225 & 172 & 42 & 176 & 15 & 42 & 1 & 132 & 238 & 95 \\
144 & 42 & 54 & 3 & 190 & 173 & 131 & 50 & 50 & 200 & 229 & 128 & 161 & 103 & 47 & 37 \\
85 & 203 & 59 & 113 & 31 & 245 & 244 & 190 & 62 & 39 & 37 & 176 & 196 & 123 & 66 & 129\text{M}
\end{matrix} \qquad (5.3)$$

The cryptanalysis carried out in this investigation strongly indicate that this cipher is a potential one and it can be applied for the transmission of text of any size and gray level/color images.

$$\begin{matrix}
2 & 95 & 7 & 118 & 130 & 53 & 132 & 123 & 10 & 29 & 82 & 113 & 143 & 155 & 35 & 59 \\
70 & 109 & 146 & 221 & 38 & 152 & 190 & 52 & 18 & 202 & 30 & 202 & 95 & 137 & 241 & 108 \\
7 & 166 & 2 & 21 & 200 & 105 & 36 & 1 & 63 & 251 & 128 & 219 & 29 & 149 & 6 & 185 \\
222 & 220 & 147 & 41 & 44 & 159 & 82 & 136 & 151 & 205 & 190 & 65 & 210 & 146 & 172 & 107 \\
30 & 241 & 25 & 80 & 94 & 28 & 65 & 18 & 241 & 112 & 50 & 238 & 98 & 5 & 204 & 255 \\
42 & 149 & 71 & 197 & 104 & 106 & 49 & 71 & 180 & 154 & 87 & 213 & 137 & 97 & 152 & 210 \\
5 & 148 & 246 & 66 & 203 & 164 & 188 & 168 & 232 & 105 & 222 & 143 & 61 & 40 & 255 & 81 \\
89 & 177 & 154 & 111 & 224 & 201 & 213 & 213 & 14 & 88 & 184 & 245 & 226 & 74 & 58 & 179 \\
148 & 70 & 31 & 251 & 7 & 1 & 126 & 100 & 11 & 249 & 244 & 73 & 240 & 58 & 212 & 23 \\
174 & 150 & 143 & 158 & 163 & 155 & 137 & 239 & 139 & 96 & 130 & 82 & 244 & 105 & 87 & 23 \\
1 & 63 & 184 & 102 & 93 & 113 & 70 & 156 & 185 & 241 & 173 & 157 & 58 & 15 & 216 & 17 \\
202 & 1 & 47 & 213 & 55 & 48 & 231 & 37 & 13 & 38 & 35 & 19 & 157 & 40 & 21 & 68 \\
61 & 71 & 255 & 44 & 109 & 131 & 140 & 98 & 182 & 14 & 95 & 176 & 239 & 38 & 129 & 113 \\
111 & 235 & 208 & 238 & 192 & 46 & 187 & 5 & 63 & 75 & 131 & 23 & 114 & 114 & 204 & 75 \\
49 & 230 & 221 & 242 & 247 & 131 & 6 & 142 & 45 & 198 & 167 & 221 & 227 & 106 & 129 & 8 \\
85 & 3 & 99 & 170 & 211 & 133 & 6 & 225 & 234 & 109 & 187 & 240 & 237 & 200 & 53 & 139
\end{matrix} \qquad (5.4)$$

and

$$\begin{matrix}
91 & 46 & 145 & 165 & 147 & 49 & 59 & 169 & 175 & 168 & 103 & 104 & 148 & 178 & 111 & 70 \\
10 & 203 & 14 & 102 & 66 & 123 & 116 & 111 & 21 & 15 & 196 & 54 & 130 & 244 & 239 & 244 \\
196 & 118 & 21 & 164 & 34 & 129 & 100 & 230 & 170 & 7 & 247 & 118 & 79 & 59 & 79 & 221 \\
38 & 189 & 221 & 142 & 11 & 39 & 142 & 255 & 168 & 49 & 78 & 150 & 157 & 183 & 101 & 161 \\
145 & 139 & 227 & 131 & 17 & 224 & 116 & 99 & 108 & 144 & 176 & 161 & 50 & 35 & 105 & 20 \\
150 & 211 & 123 & 240 & 174 & 55 & 101 & 210 & 141 & 87 & 83 & 246 & 46 & 70 & 53 & 46 \\
108 & 46 & 80 & 112 & 172 & 232 & 228 & 69 & 97 & 232 & 166 & 102 & 70 & 63 & 94 & 18 \\
18 & 214 & 79 & 151 & 79 & 250 & 10 & 116 & 81 & 115 & 228 & 77 & 121 & 12 & 153 & 167 \\
131 & 133 & 132 & 246 & 53 & 94 & 132 & 39 & 151 & 183 & 207 & 36 & 194 & 222 & 227 & 70 \\
141 & 193 & 91 & 210 & 120 & 146 & 152 & 224 & 202 & 110 & 34 & 0 & 73 & 176 & 4 & 0 \\
198 & 2 & 215 & 20 & 141 & 203 & 183 & 92 & 214 & 217 & 37 & 140 & 141 & 161 & 211 & 248 \\
25 & 70 & 49 & 234 & 95 & 31 & 25 & 99 & 200 & 248 & 251 & 243 & 15 & 149 & 206 & 78 \\
116 & 135 & 103 & 157 & 37 & 64 & 242 & 116 & 246 & 219 & 17 & 71 & 249 & 157 & 127 & 34 \\
17 & 148 & 51 & 32 & 121 & 45 & 163 & 192 & 14 & 166 & 62 & 211 & 64 & 156 & 40 & 50 \\
220 & 210 & 244 & 208 & 41 & 113 & 132 & 254 & 115 & 174 & 22 & 231 & 196 & 188 & 67 & 126 \\
121 & 96 & 4 & 234 & 70 & 102 & 186 & 145 & 133 & 69 & 222 & 158 & 239 & 30 & 75 & 146
\end{matrix} \tag{5.5}$$

REFERENCES

[1] Dr.V.U.K.Sastry, K.Shirisha, "A Novel Block Cipher Involving a Key Bunch Matrix and a Key-based Permutation and Substitution", in International Journal of Advanced Computer Science and Applications(IJACSA), Vol. 3, No. 12, Jan 2012, pp.16-122.

[2] 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 87 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 9 research papers in International Journals. She stood University topper in the M.Tech.(CSE).