IJCSIS Vol. 7, No. 1, January 2010
ISSN 1947-5500

International Journal of Computer Science & Information Security

IJCSIS Editorial

Message from Managing Editor

IJCSIS Volume 7 No. 1 January 2010 issue presents high-quality e-publications (with acceptance rate of ~ 28%) that deals will all aspects of computer science. Paper selection undergoes journal-style peer review process. The editorial board and technical review committee contain some of the most renowned specialists in their areas of expertise. In this way, both the high standard of the published papers as well as a broad spectrum of publications is guaranteed.

In this issue, you will find high-quality contributions from researchers in the diverse area of computer science, networking, information retrieval, emerging technologies and information security. With a continuing open-access policy, we are pleased to disseminate this collection of remarkable computer science research works with our readers.

Special thanks to our reviewers and technical sponsors for their valuable service.

Available at http://sites.google.com/site/ijcsis/

IJCSIS Vol. 7, No. 1,

January 2010 Edition

ISSN 1947-5500

Indexed by (among others):

IJCSIS EDITORIAL BOARD

TABLE OF CONTENTS

Md. Rafiqul Islam, A.W. Naji, A. A. Zaidan and B. B. Zaidan
Department of Electrical and Computer Engineering, Faculty of Engineering, International Islamic University Malaysia (IIUM), P.O. Box 10, 50728 Kuala Lumpur, Malaysia

45. Paper 07011064: An innovative platform to improve the performance of exact-string-matching algorithms (pp. 280-283)

Mosleh M. Abu-Alhaj[1], M. Halaiyqah[2], Muhannad A. Abu-Hashem[2], Adnan A. Hnaif[1], O. Abouabdalla[1], and Ahmed M. Manasrah.
[1]: National Advanced IPv6 Center of Excellence, [2]: Computer Science
University Sains Malaysia, Penang Malaysia

46. Paper 18011012: A MAC Layer Based Defense Architecture for Reduction-of-Quality (RoQ) Attacks in Wireless LAN (pp. 284-291)

Jatinder Singh, Director, Universal Institute of Engg. & Tech. Lalru-CHD (India)
Dr. Savita Gupta, Prof. Deptt. Of Computer Engg., UIET, PunjabUniversity, CHD, India
Dr. Lakhwinder Kaur, Reader, UCOE, Punjabi University, Patiala, India

47. Paper 06011052: Application of k-Means Clustering algorithm for prediction of Students' Academic Performance (pp. 292-295)

Oyelade, O. J, Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria.
Oladipupo, O. O, Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria.
Obagbuwa, I. C, Department of Computer Science, Lagos State University, Lagos, Nigeria.

SCENARIO BASED WORM TRACE PATTERN IDENTIFICATION TECHNIQUE

Siti Rahayu S., Robiah Y., Shahrin S., Mohd Zaki M., Irda R., Faizal M. A.
Faculty of Information and Communication Technology
Univeristi Teknikal Malaysia Melaka,
Durian Tunggal, Melaka,
Malaysia
.

Abstract—**The number of malware variants is growing tremendously and the study of malware attacks on the Internet is still a demanding research domain. In this research, various logs from different OSI layer are explore to identify the traces leave on the attacker and victim logs, and the attack worm trace pattern are establish in order to reveal true attacker or victim. For the purpose of this paper, it will only concentrate on cybercrime that caused by malware network intrusion and used the traditional worm namely blaster worm variants. This research creates the concept of trace pattern by fusing the attacker's and victim's perspective. Therefore, the objective of this paper is to propose on attacker's, victim's and multi-step (attacker/victim)'s trace patterns by combining both perspectives. These three proposed worm trace patterns can be extended into research areas in alert correlation and computer forensic investigation.**

Keywords— **trace pattern, attack pattern, log**

I. INTRODUCTION

Nowadays the numbers of cases of internet threat causes by malware have been tremendously increased. Malware that consist of Trojan, virus and worm had threatened the internet user and causes billion of losses to the internet users around the world. The exponential growth of malware variants as reported by AV-test.org [1] is quite alarming. This can be seen from the higher growth rate of the malware collection shown in the graph of the new malware samples over the last several years as depicted in Fig.1.

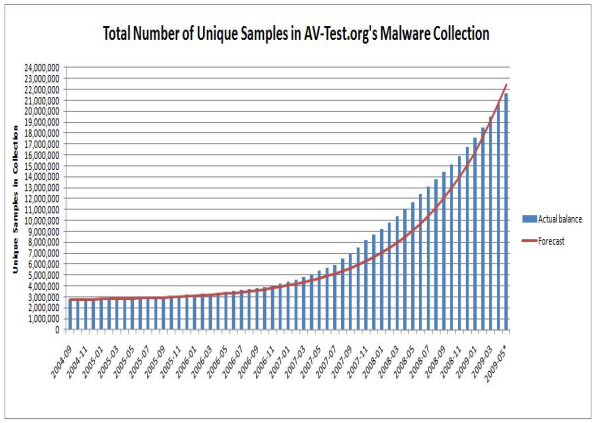


Fig. 1 The continuous growth of the number of malware [1]

In the report, the numbers of malware is not cumulative where they represent only the new variants in the time frame specified without including the previously identified ones. This does not mean that there are only unique pieces of malware as there are also many variants of the same pieces of malware. Variants are often created to defeat the security tools, for instance a worm can mutate to a different variants, sometimes in only one hour [2]. Thus make it difficult for security tool to detect the threat.

Due to this reason the study of attacks on the internet has becoming an important research domain. The study on internet attack is very crucial, especially in developing an effective security tool to defend the internet user from the attack threat. Collecting data related to Internet threats has become a relatively common task for security researchers and from the data the researcher can investigate the attack pattern in order to find the root cause and effect of an attack. Attack pattern can also be used as a guide to the investigator for collecting and tracing the evidence in forensic field [3]. According to [4], the anatomy of attack consists of attacker and victim perspective.

To address this challenge, this paper propose on attacker's, victim's and multi-step (attacker/victim)'s trace patterns by collaborate both attacker and victim perspectives. This research explore the various logs from different OSI layer to identify the traces leave on the attacker and victim logs, and establish the attack trace pattern in order to reveal true attacker or victim. The research only focuses on cybercrime that caused by malware network intrusion specifically traditional worm namely blaster worm variants.

II. RELATED WORK

A. *Blaster Worm*

Generally malware consists of virus, worm and trojan horse [5]. For the purpose of this research, the researchers focuses on one types of worm that described by [6], which is traditional worm specifically blaster worm.

Blaster uses a public exploit for the RPC DCOM vulnerability in order to obtain a system-level command shell on its victims [7]. The worm start searching for IP addresses at a time for hosts with listening TCP port 135 open by sending a connection attempt to each one simultaneously. Once found, it tries to figure out the windows version and then sends the RPC DCOM exploit that binds a system level shell to TCP port 4444. Once the target is successfully compromised, the worm transmits the msblast.exe executable via TFTP on UDP port 69 to infect the host where the payloads used in the public DCOM exploit, as well as the

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

TFTP functionality, are both encapsulated within msblast.exe. Once the executable has been transferred, or after 20 seconds have elapsed, the TFTP server is shut down and the worm then issues further commands to the victim to execute msblast.exe [8]. Assuming the executable was downloaded successfully, the propagation cycle then begins again from the newly infected host, while the infecting instance of the worm continues iterating through IP addresses.

B. *Trace Pattern*

Trace is described as a process of finding or discovering the origin or cause of certain scenario and pattern is defined as a regular way in which certain scenario happened [9]. Trace pattern is essential in assisting the investigators tracing the evidence found at crime scenes. In the computer crime perspective, it can be found in any digital devices. These traces consist in a variety of data records their activities such as login and logout to the system, visit of pages, documents accesses, items created and affiliation to groups. Traces data are typically stored in log files and normally takes on several selected attributes such as port, action, protocol, source IP address and destination IP address.

The trace data can be used to identify a victim or attacker by analyzing the attack pattern which is represented in the form of trace pattern can helps determine how a crime is being committed. Attack pattern is type of pattern that is specified from attacker perspective. The pattern describes how an attack is performed, enumerates the security patterns that can be applied to defeat the attack, and describes how to trace the attack once it has occurred [10].

An attack pattern provides a systematic description of the attack goals and attack strategies for defending against and tracing the attack. Hence, attack patterns can guide forensic investigators in searching the evidence and the patterns can serve as a structured method for obtaining and representing relevant network forensic information. This also helps the forensic investigator at the data collection phase that requires the investigator to determine and identifying all the components to be collected, deciding the priority of the data, finding the location of the components and collecting data from each of the component during the investigation process [11].

Various descriptions provided by several researchers to describe the term attack pattern. According to [12], they use the term attack pattern to describe the steps in a generic attack. Meanwhile, [13] describe the term attack pattern as the attack steps, attack goal, pre-conditions and post-conditions of an attack. [14] describe an attack pattern as the approach used by attackers to generate an exploit against software in which it is a general framework for carrying out a particular type of attack, such as a method for exploiting a buffer overflow or an interposition attack that leverages certain kinds of architectural weaknesses.

Although there are different descriptions provided by several researchers, they have the same idea and concept that the attack pattern is very important to provide a way to protect them from any potential attack. For example, software developers use attack pattern to learn about how their software may be attacked. Armed with knowledge about potential attacks, developers can take steps to mitigate the impact of these attacks. Similarly, network administrator or network engineers use attack pattern to study on how the potential attacker attack their network in order to detect and block any vulnerabilities in their network.

The study from [12][13][14] discussed the concept of attack patterns as a mechanism to capture and communicate at the attacker's perspective that shows the common methods for exploiting software, system or network while [10] and [11] discussed on the attack pattern on how the attack is performed, the attack goals, how to defences against the attack and how to trace once it has occurred.

All of the researchers are only focusing on the attacker's perspective while victim's perspective is omitted. Therefore, this research proposed the trace patterns by focusing on the attacker's, victim's and attacker/victim's (multi-step) perspectives to provide clear view on how the attacker performed the attack and what is the impact caused by the attack.

Multi-step perspective proposed in this research is motivated based on the study by [15] in order to reveal the true attacker or victim. A multi-step attack is a sequence of attack steps that an attacker has performed, where each step of the attack is dependent on the successful completion of the previous step. These attack steps are the scan followed by the break-in to the host and the tool-installation, and finally an internal scan originating from the compromised host [16].

In the next section, researchers present the experimental approach used in this research to gather and analyse logs for designing the proposed worm trace pattern.

III. EXPERIMENTAL APPROACH

This experimental approach used four phases: Network Environment Setup, Attack Activation, Trace Pattern Log Collection and Trace Pattern Log Analysis. Further details on these phases are explained as follows and depicted in Fig. 2.

Fig. 2 Trace Pattern Experimental Approach Framework

A. *Network Environment Setup*

The network environment setup use in this experiment is following the network simulation setup done by the MIT Lincoln Lab [17] and it only consists of Centos and Windows XP operating systems to suit our experiment's environment. The network design is shown in Fig. 3.

This network design consists of two switches, one router, three servers for Intrusion Detection System (IDS Arowana

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

and IDS Sepat) and Network Time Protocol (NTP) run on *Centos 4.0*, seven victims and one attacker run on *Windows XP*.

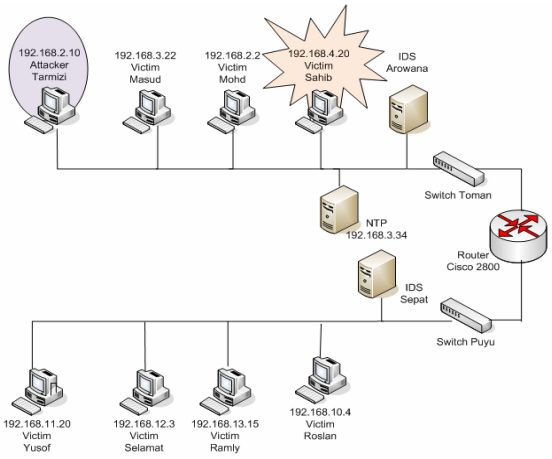


Fig. 3 Network Setup Environment for Blaster Trace Pattern

In this experiment, host 192.168.2.10 (*Tarmizi*) is Attacker, host 192.168.4.20 (*Sahib*) and 192.168.11.20 (*Yusof*) are the selected Victims for this research. The log files that are expected to be analyzed are *personal firewall log*, *security log*, *system log* and *application log*. Each log are generated by host level device and one log files by network level device (*alert log* by IDS). *Wireshark* were installed in each host and *tcpdump* script is activated in IDS to capture the whole network traffic. The *Wireshark* and *tcpdump* script were used to verify the traffic between particular host and other device.

B. *Attack Activation*

The attacker machine, *Tarmizi* is selected to launch the Blaster variant attack. This experiment runs for one hour without any human interruption in order to obtain the attacker and victim logs.

C. *Trace Pattern Log Collection*

Each machine generated *personal firewall log, security log, application log, system log* and *wireshark log*. The IDS machine generates *alert log* and *tcpdump log*. The trace pattern logs are collected at each victim and attacker machine. The *wireshark* and *tcpdump* files are used to verify the traffic between particular host and other device.

D. *Trace Pattern Log Analysis*

In this trace pattern log analysis process the researchers analyze the logs generated by the attacker and victim machine. The objective of the trace pattern log analysis is to identify the victim and attacker trace pattern by observing the traces leave on the selected logs. The output from the analysis is used to construct worm trace pattern as proposed in Section V.

IV. Analysis and Findings

The researchers have collected all logs generated during the experiment and the attack scenario is identified. Based on the scenario, various logs from attacker's host, victim's host, and network are analyzed.

A. *Attack Scenario*

The attack scenario as depicted in Fig. 4 is derived based on the log analysis in the experimental approach framework in section III.

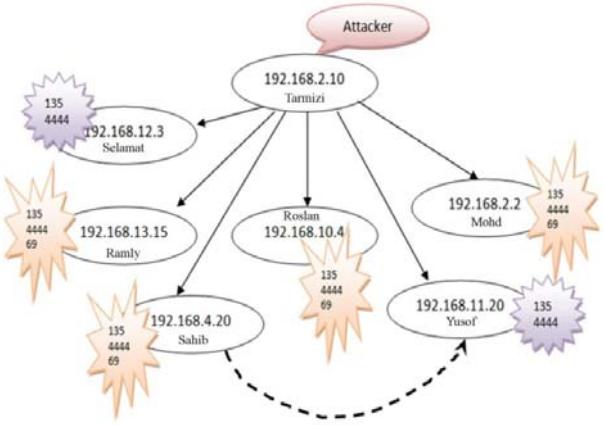


Fig. 4 Blaster Attack Scenario

Based on the analysis, the worm attack is activated in *Tarmizi* and successfully exploited all hosts that mark with *135*, *4444* and *69* except for host *Yusof* and *Selamat*. These hosts was mark with *135* and *4444* and indicated that the attacker is already open the backdoor but unable to transfer the malicious codes through port *69*. Then, one of the infected hosts that are *Sahib* has organized attack on host *Yusof*.

B. *Trace Pattern Analysis*

The purpose of trace pattern analysis is to construct the victim's, attacker's and multi-step (victim/attacker)'s trace pattern by observing the traces leave on the selected logs. The various logs involve in this analysis are host logs: *personal firewall log, security log, system log, application log* and network logs: *alert log* by IDS. Based on the attack scenario described previously, the logs are further analyzed to extract the trace pattern generated by the attack from the attacker's and victim's machine by tracing the fingerprints emerged in the logs.

In this analysis, the researchers have selected *Sahib* and *Yusof* as victims; and *Tarmizi* as attacker as shown in Fig. 3. The tracing tasks involve the log from devices at host and network level and initially started at victim's logs followed by attacker's logs: *personal firewall log, security log, system log* and *application log*. The network logs are used to complement the finding from the host level tracing tasks. The details of the analysis for victim, attacker and multi-step attack (attacker/victim) trace pattern are further elaborated in the next sub-section.

1) *Victim's Traces*

The victim's data traces are extracted from the logs at the victim's host and network log. The summary of the data traces are depicted in Fig. 5 and the evidences are found in *personal firewall log, security log, system log* and *application log* and the details of the traces are discussed.

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

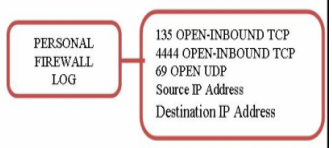
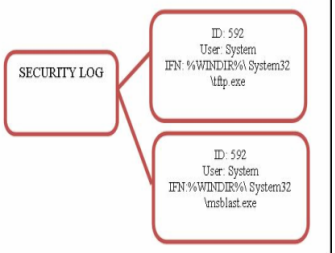
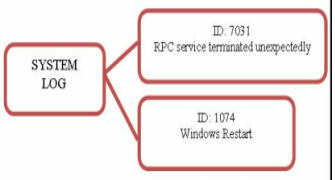
Level	Evidence/Log	Category	Trace Pattern	Extracted Data	Conceptual Diagram
Host	Personal Firewall Log	Scan and Exploit	**Action, Protocol and Destination Port:** OPEN-INBOUND 135 TCP OPEN-INBOUND 4444 TCP OPEN 69 UDP **Source IP Address and Destination IP Address:** Refer to Table 1	2009-09-07 14:41:09 OPEN-INBOUND TCP 192.168.2.10 192.168.4.20 3993 135 - - - - - - - - 2009-09-07 14:41:12 OPEN-INBOUND TCP 192.168.2.10 192.168.4.20 4002 4444 - - - - - - - 2009-09-07 14:41:12 OPEN UDP 192.168.4.20 192.168.2.10 3027 69 - - - - - - - -	
	Security Log	Impact / Effect	**Event ID:** 592 **User ID:** System **Image File Name (IFN):** • %WINDIR%\ System32 \tftp.exe • %WINDIR%\ System32 \msblast.exe	09/07/2009 14:41:12 Security Success Audit Detailed Tracking 592 NT AUTHORITY\SYSTEM SAHIB **"A new process has been created:** New Process ID: 1016 Image File Name: C:\WINDOWS\system32\tftp.exe Creator Process ID: 1228 User Name: SAHIB$ Domain: WORKGROUP Logon ID: (0x0,0x3E7) 09/07/2009 14:41:25 Security Success Audit Detailed Tracking 592 NT AUTHORITY\SYSTEM SAHIB **"A new process has been created:** New Process ID: 1752 Image File Name: C:\WINDOWS\system32\msblast.exe Creator Process ID:1228 User Name: SAHIB$ Domain: WORKGROUP Logon ID: (0x0,0x3E7)	
	System Log		**Event ID:** 7031 1074 **Image File Name (IFN):** RPC service terminated unexpectedly Windows Restart	09/07/2009 14:41:29Service Control Manager Error None 7031 N/A SAHIB **The Remote Procedure Call (RPC) service terminated unexpectedly.** It has done this 1 time(s). The following corrective action will be taken in 60000 milliseconds: Reboot the machine. 09/07/2009 14:41:29USER32 Information None 1074 NT AUTHORITY\SYSTEM SAHIB The process winlogon.exe has initiated the restart of SAHIB for the following reason: No title for this reason could be found Minor Reason: 0xff Shutdown Type: reboot Comment: **Windows must now restart because the Remote Procedure Call (RPC) service terminated unexpectedly**	
Network	Alert IDS Log	Activity	**Message:** TFTP Get	[**] [1:1444:3] TFTP Get [**] [Classification: Potentially Bad Traffic] [Priority: 2]	
		Alarm	**Source IP:** SRC_IP: Victim **Destination IP:** Dest_IP: Attacker **Destination Port:** 69	09/07-14:41:03.846382 192.168.4.20:3027 -> 192.168.2.10:69 UDP TTL:128 TOS:0x0 ID:337 IpLen:20 DgmLen:48 Len: 20	

Fig. 5 Summary of Blaster Traces on Victim's Log

In *Personal Firewall log*, ports 135 TCP, 4444 TCP and 69 UDP are considered as part of this victim's trace pattern due to the information gain from [7], [18] and [19]. These ports are known as vulnerable ports that can be used by malicious codes to exploit the system-level command shell on its victims. These ports also provide an inter-process communication mechanism that allows programs running on one host to execute code on remote hosts [20].

The source IP address of OPEN-INBOUND TCP indicates the remote host and the source IP address of OPEN UDP indicates the local host. While the destination IP address of OPEN-INBOUND TCP indicates the local host and the destination IP address of OPEN UDP indicates the remote host. The summarized of the IP address dependency is depicted in TABLE 1.

TABLE 1: IP ADDRESS DEPENDENCY

Action	Source IP Address	Destination IP Address
OPEN-INBOUND	Remote Host	Local Host
OPEN	Local Host	Remote Host

The patterns of the communication between Source IP address and Destination IP address indicate that the local host has permitted the TFTP service and the incoming traffic from the remote host. All communication activity is done by vulnerable ports open exploitation.

In *Security log*, the traces data from the security log shows that there is a new process created by system proves by the existence of *event id 592*. It has initiated the TFTP service used to download and upload the blaster worm code and execute the remote blaster worm code (msblast.exe). This activity is logged in *Image File Name* as *%WINDIR%\System32\tftp.exe* and *%WINDIR%\System32\msblast.exe*.

System log shows the traces data of the Blaster-infected machine stops its TFTP daemon after a transmission or after 20 seconds of TFTP inactivity by showing the new process created on event id 7031 and 1074 that indicates the RPC service terminated unexpectedly and windows restart respectively.

The *alert IDS log* shows that there is an activity called TFTP Get is occurred on port 69 UDP where the source IP

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

address from alert IDS log is similar as the destination IP address in the victim's firewall log and vice versa. These traces identify that there is a pattern exists on how the blaster worm initiates the client to download the worm code. The Source IP address is the victim and the Destination IP address is the attacker.

Action OPEN-INBOUND, OPEN-INBOUND and OPEN are known as a complete sequence of communication of blaster worm to gain access and upload the malicious codes to be exploited [21] where the OPEN action shows that an outbound session was opened to a remote host and the OPEN-INBOUND action shows that an inbound session was opened to the local host

2) Attacker's Traces

The attacker's data traces are extracted from the logs at the attacker's host and network log. The summary of the data traces on the attacker's and network logs are illustrated in Fig. 6 and the evidences are found in *personal firewall log*, *security log*, *system log* and *application log* and the details of the traces of the attacker's logs are discussed as following.

The traces data leaved in attacker's *Personal Firewall Log* shown the vulnerable ports that are used by the attacker to exploit the system-level command shell on its victims. The pattern of the traces data are OPEN TCP 135, OPEN TCP 4444 and OPEN-INBOUND UDP 69 as referred to [7], [18] and [19].

The source IP address of OPEN TCP indicates the local host and the destination IP address action indicates the remote host. While, the source IP address of OPEN-INBOUND UDP indicates the remote host and the destination IP address indicates the local host.

The patterns of source IP address and destination IP address indicate that the local host is opened an outbound session to the remote host which allow the local host transmit the payload (worm codes) to the remote host by exploiting the vulnerable ports open. The traces data from the *security log* shows that there is a new process created (Event ID: 592) that shows the blaster worm is activated based on the trace shows on the image file name.

In the *alert IDS log*, (Portscan) TCP Portsweep presents the pattern of scanning activity which shows the behavior of traditional worm attack in general and blaster worm attack in specific [22]. Therefore, this trace discovers that the owner of the source IP address is a potential attacker who launched the worm.

3) Multi-step (Attacker/Victim) Trace Pattern

Based on the extracted data from the logs at the victim's host and network log, the multi-step (Attacker/Victim)'s traces data is identified. The summary of the data traces on the multi-step and network logs are represented in Fig. 7 and the evidences are found in *personal firewall log*, *security log*, *system log* and *application log* and the details of the traces of the multi-step's logs are discussed.

There are two different patterns existing in *personal firewall log* that discover attacker and victim traces as shown in TABLE 2.

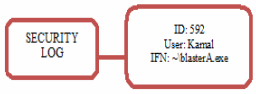
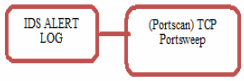
Level	Evidence/ Log	Category	Trace Pattern	Extracted Data	Conceptual Diagram
Host	Personal Firewall Log	Scan and Exploit	**Action, Protocol and Destination Port:** OPEN 135 TCP OPEN 4444 TCP OPEN-INBOUND 69 UDP **Source IP Address and Destination IP Address:** Refer to Table 1	2009-09-07 14:41:09 **OPEN TCP** 192.168.2.10 192.168.4.20 3993 **135** - - - - - - - - 2009-09-07 14:41:11 **OPEN TCP** 192.168.2.10 192.168.4.20 4002 **4444** - - - - - - - - 2009-09-07 14:41:11 **OPEN-INBOUND UDP** 192.168.4.20 192.168.2.10 3027 **69** - - - - - - - -	
	Security Log	Symptom	**Event ID:** 592 **Image File Name (IFN):** ~\blasterA.exe	09/07/2009 14:40:01 Security Success Audit Detailed Tracking **592** Kamal TARMIZI "A new process has been created: New Process ID: 1408 Image File Name: C:\Documents and Settings\Kamal\Desktop**blasterA.exe** Creator Process ID: 1492 User Name: Kamal Domain: TARMIZI Logon ID: (0x0,0x2273F)	
Network	Alert IDS Log	Activity	**Message:** (Portscan) TCP Portsweep	[**] [122:3:0] **(portscan) TCP Portsweep** [**] [Priority: 3]	
		Alarm	**Source IP:** SRC_IP: Attacker	09/07-14:44:18.729996 **192.168.2.10** -> **192.168.11.1** PROTO:255 TTL:0 TOS:0x0 ID:3307 IpLen:20 DgmLen:166	

Fig. 6 Summary of Blaster Traces on Attacker's Log

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

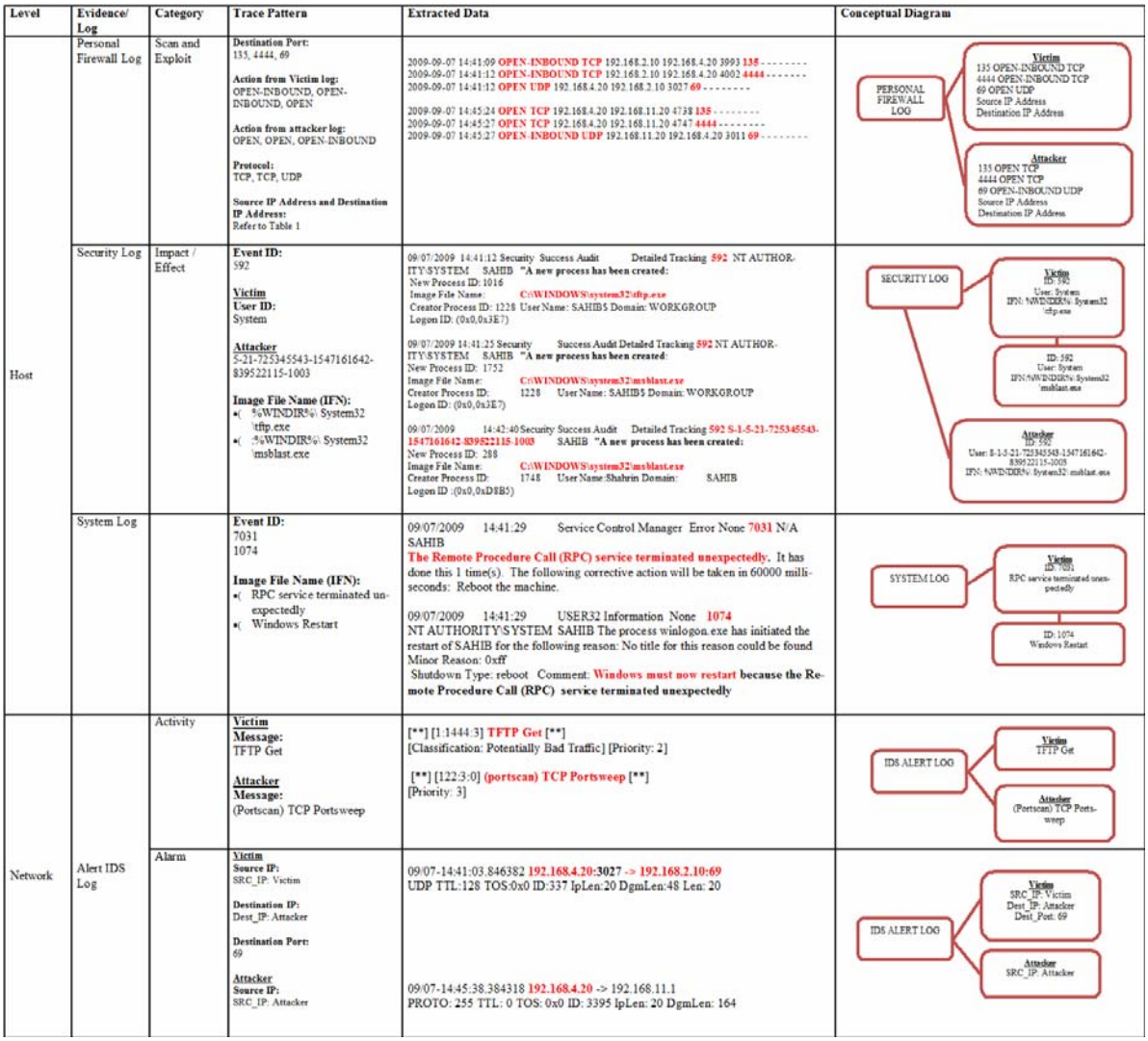


Fig. 7: Blaster Traces on Multi-step (Victim/Attacker) Log

TABLE 2: TRACES DATA ON PERSONAL FIREWALL LOG

Victim	Attacker
Action, Protocol and Destination port: OPEN-INBOUND TCP 135, OPEN-INBOUND TCP 4444 and OPEN UDP 69 - These ports are known as vulnerable ports that provide hole and opportunities to the attacker to exploit in order to gain access to the system.	**Action, Protocol and Destination port:** OPEN TCP 135, OPEN TCP 4444and OPEN -INBOUND UDP 69 - These ports are known as vulnerable ports that can be used by the attacker to exploit the system-level command shell on its victims.
Source IP Address and Destination IP Address: OPEN-INBOUND TCP - the source IP address indicates the remote host and the destination IP address indicates the local host. OPEN UDP - the source IP address indicates the local host and the destination IP address indicates the remote host.	**Source IP Address and Destination IP Address:** OPEN TCP - the source IP address indicates the local host and the destination IP address indicates the remote host. OPEN-INBOUND UDP - the source IP address indicates the remote host and the destination IP address indicates the local host.

From the victim perspective *(OPEN-INBOUND TCP 135, OPEN-INBOUND TCP 4444 and OPEN UDP 69)*, the patterns shows that the source IP address and destination IP address indicate that the local host is permitted the *TFTP* service and the incoming traffic from the remote host.

While, from the attacker perspective *(OPEN TCP 135, OPEN TCP 4444* and *OPEN -INBOUND UDP 69)*, the patterns of source IP address and destination IP address indicate that the local host is opened an outbound session to the remote host which allow the local host transmit the payload (worm codes) to the remote host.

All communication activity is done by vulnerable ports open exploitation. Therefore, the traces data found are significant to the multi-step attack (victim/attacker) where this host was infected (act as victim) and as long as the computer was infected with the worm code *(msblast)*, it (act as attacker) continued to generate traffic to attempt to infect other vulnerable computers [23].

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

The traces data in TABLE 3 are taken from the *security log* in Fig. 9 and it shows that there is a new process created by system which initiates the *TFTP* service. This service is used to receive and sent the blaster worm code and execute the blaster worm code (*msblast.exe*) remotely.

TABLE 3: TRACES DATA ON SECURITY LOG

Victim	Attacker
Event ID: 592	**Event ID**: 592
User ID: System	**User ID:** S-1-5-21-725345543-1547161642-839522115-1003
Image File Name: %WINDIR%\System32\ tftp.exe, %WINDIR%\System32\msblast.exe	**Image File Name** %WINDIR%\System32\msblast.exe

The pattern in TABLE 3 has indicates that this host is a victim of blaster worm attack. Another traces data found is a new process created (Event ID: 592) by unidentified user (S-1-5-21-725345543-1547161642-839522115-1003) which executed the *msblast* as shown on the image file name. It identify that this host was infected previously and automatically attempt to transfer the worm code by generating traffic to exploit other vulnerable computers.

System log shows the traces data of the Blaster-infected machine stops by showing the new process created on event id 7031 and 1074 that indicates the RPC service terminated unexpectedly and windows restart respectively. This pattern is significant with the victim pattern in which if the host is infected by blaster worm, the RPC service is terminated unexpectedly by exploiting the RPC DCOM and force the windows restart.

As depicted in TABLE 4, the *alert IDS log* shows that there are traces found on TFTP Get and (Portscan) TCP Portsweep activities for victim and attacker respectively.

The *TFTP Get* activity on port *69* trace indicates that there is a pattern exists on how the blaster worm initiates the client to download the worm code. The source IP address is the victim and the destination IP address is the attacker. On the other hand, the *(Portscan) TCP Portsweep* trace proves that there is scanning activity on the vulnerable open port. This trace indicates that the source IP address is the attacker who activated the worm. Both traces found in TABLE 4 are significant to the pattern that found in victim and attacker as depicted in Fig. 8 and Fig. 9 respectively.

TABLE 4: TRACES DATA ON ALERT IDS LOG

Victim	Attacker
Message: TFTP Get	**Message**: (Portscan) TCP Portsweep
Source IP Address and Destination IP Address: Source IP address indicates the victim and the Destination IP address indicates the attacker	**Source IP Address**: Source IP address indicates the attacker
Destination Port: 69	

Based on the analysis, the researchers have identified the significant attributes from the victim, attacker and multi-step traces data. These findings are further use to construct the proposed worm trace pattern.

V. PROPOSED WORM TRACE PATTERN

This research proposed the worm trace pattern based on victim, attacker and multi-step point of view. The following section describes the details.

A. *Victim's Trace Pattern*

Victim's trace pattern is useful for forensic in order to provide clear view on how the victim attacked by the potential attacker. According to the analysis and findings from Fig. 5, the overall Blaster victim's trace pattern is summarized in Fig.8.

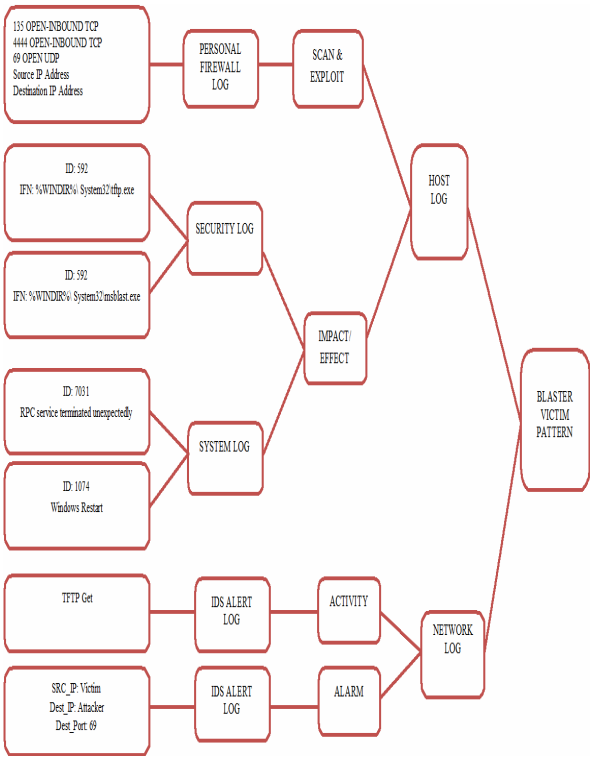


Fig.8 Proposed Blaster Victim's Trace Pattern

In Fig. 8, the traces data indicated the blaster worm pattern at the victim's host used port 135 TCP to permit the scanning and transmitting RPC DCOM exploit codes from remote host which launch the windows shell to initiate worm code download used port 4444 TCP.

Then it launched the TFTP client service using port 69 for downloading the worm code. The traces on TFTP Get on port 69 UDP also found in the network log that supports all the traces found on the host log.

B. *Attacker's Trace Pattern*

Attacker's trace pattern provides a systematic description of the attack goals and attack strategies for defending against and tracing the attack. This pattern is useful to guide forensic investigators in searching the evidence and provide a

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

structured method for obtaining and representing relevant network forensic information.

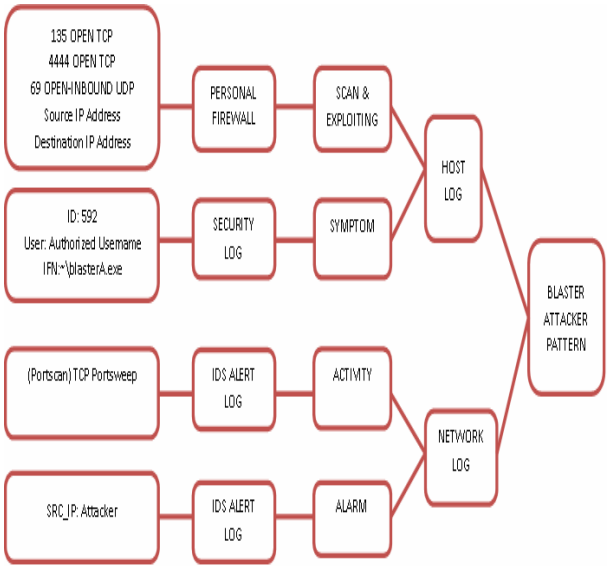


Fig. 9 Proposed Blaster Attacker's Trace Pattern

The overall Blaster attacker's trace pattern depicted in Fig. 9 indicate the blaster worm pattern at the attacker's host used port *135 TCP* to allow the local host scan and transmit *RPC DCOM* exploit codes to the remote host which launch the windows shell to initiate worm code download used port *4444 TCP*. Then it launched the *TFTP* client service using port *69* to permit the client (remote host) download the worm code from the local host. The activity of *TCP Portsweep* trace also existed in the network log that supports all the traces found on the host log.

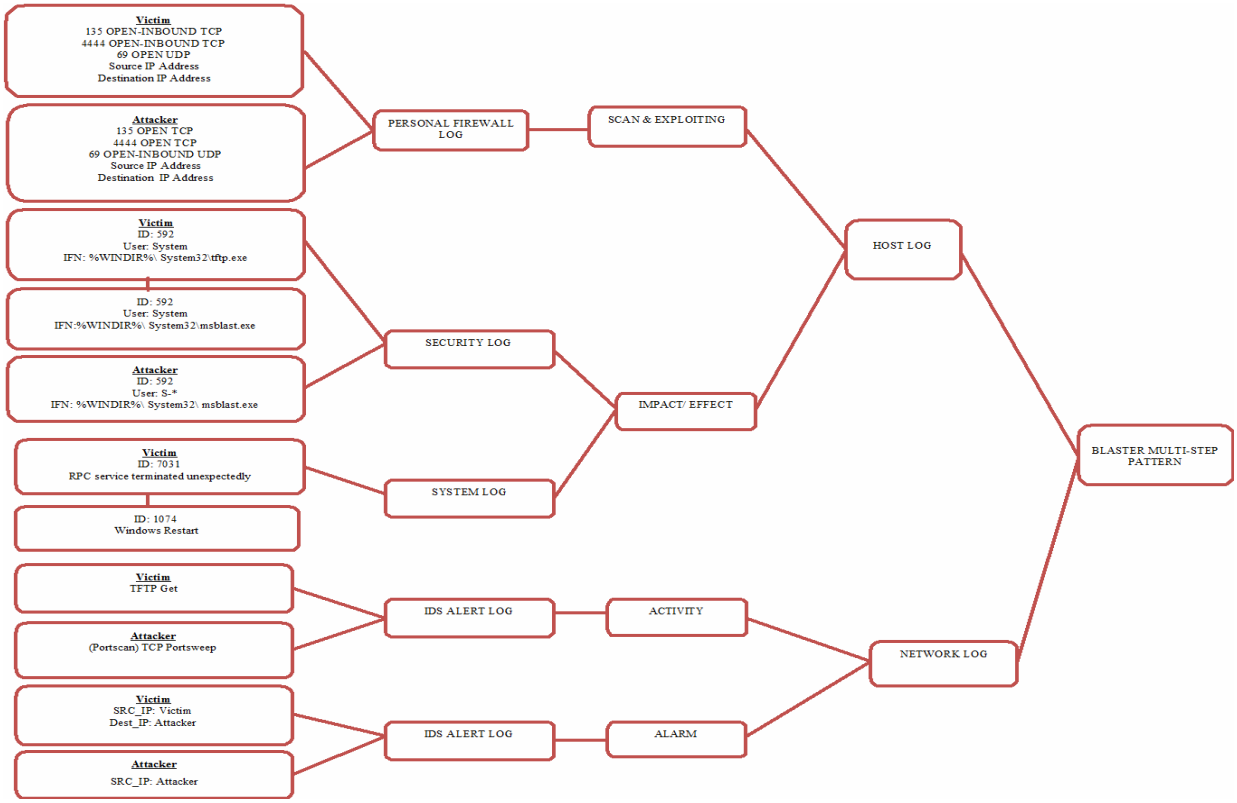
C. Multi-step (Attacker/Victim) Trace Pattern

Multi-step's trace pattern is used as a guide for forensic investigators to reveal and prove the true attacker or victim. This trace pattern is a combination of victim's and attacker's trace pattern in which the traces data is extracted from a log for the same host.

The traces data on multi-step at the host's logs from victim/attacker perspective illustrated in Fig. 10 indicate that the blaster worm used port 135 TCP to permit the scanning activity and it is supported by the traces found in network logs that show *(Portscan) TCP Portsweep* activities.

The worm then transmit RPC DCOM exploit codes from remote host which launch the windows shell to initiate downloading the worm code using port 4444 TCP and it launched the TFTP client service on port 69. This worm activity is shown by the traces found in network logs that confirm the existence of *TFTP Get on port 69 UDP* activities. Once the host is infected (act as victim), it's (act as attacker) then generate traffic; attempt to infect other vulnerable hosts.

The source IP address from host log indicates that the remote host is the victim and the destination IP address which is the local host is the attacker. Hence, multi-step (victim/attacker) trace pattern could identify the true victim or attacker.



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

Fig. 10 Proposed Multi-step (Victim/Attacker) Trace Pattern

VI. CONCLUSIONS AND FUTURE WORKS

Trace pattern of an attack in an attack scenario is constructed by analyzing various logs from heterogeneous devices in victim, attacker and victim/attacker (multi-step) perspectives. These trace patterns offer a systematic description of the impact of the attack, the attack goals and attack steps along with strategies for defending against and tracing the attack. For example, personal firewall logs provide information on how the attacker entered the network and how the exploits were performed; meanwhile event logging such as security log, system log and application log enables network administrators to collect important information such as date, time and result of each action during the setup and execution of an attack. Therefore, the propose victim, attacker and multi-step (victim/attacker) trace patterns in this paper can be extended into research areas in alert correlation and computer forensic investigation.

REFERENCES

[1] Seltzer, L. (2009). The Growth of Malware: Up, Up, and Away [Electronic Version]. Retrieved 24 Nov 2009 from http://blogs.pcmag.com/securitywatch/2009/07/the_growth_of_malware_up_up_an.php.

[2] Gadelrab, M., Abou El Kalam, A., & Deswarte, Y. (2008). *Execution Patterns in Automatic Malware and Human-Centric Attacks.* Paper presented at the Proceeding of Seventh IEEE International Symposium on Network Computing and Applications.

[3] Olivier, T., & Marc, D. (2008). A framework for attack patterns' discovery in honeynet data. *Journal of Digital Investigation, 5*, 128-139.

[4] McHugh, J. (2001). Intrusion and intrusion detection. *International Journal of Information Security, 1*, 14-35.

[5] Karresand, M. (2003). *A proposed taxonomy of software weapons* (No. FOI-R-0840-SE): FOI-Swedish Defence Research Agency.

[6] Lazarevic, A., Kumar, V., & Srivastava, J. (2005). Managing Cyber Threats (Pg 19-78). On *Massive Computing*: Springer US.

[7] Microsoft. (2003). Virus alert about the Blaster worm and its variants [Electronic Version]. Retrieved 23 July 2009 from http://support.microsoft.com/kb/826955.

[8] Bailey, M., Cooke, E., Jahanian, F., Watson, D., & Nazario, J. (2005). The Blaster Worm: Then and Now. *IEEE Computer Society*.

[9] Hornby, A. S. (2005). *Oxford Advanced Learner's Dictionary* (7 ed.). New York: Oxford University Press

[10] Fernandez, E., Pelaez, J., & Larrondo-Petrie, M. (2007). Attack Patterns: A New Forensic and Design Tool. *IFIP International Federation for Information Processing, 242*, 345-357.

[11] Kent, K., Chevalier, S., Grance, T., & Dang., H. (2006). Guide to Integrating Forensic Techniques into Incident Response, NIST Special Publication 800-86.

[12] Hoglund, G., & McGraw, G. (2004). *Exploiting Software: How to Break Code*. Boston, Massachusetts: Addison-Wesley/Pearson.

[13] Moore, A., Ellison, R., & Linger, R. (2001). *Attack modeling for information security and survability, Technical Note CMU/SEI-2001-TN-001*. Pittsburgh, Pennsylvania: Software Engineering Institute, Carnegie Mellon University

[14] Sean, B., & Amit, S. (2006). Introduction to Attack Patterns [Electronic Version]. Retrieved 19 Nov 2009.

[15] Liu, Z., Wang, C., & Chen, S. (2008). Correlating Multi-Step Attack and Constructing Attack Scenarios Based on Attack Pattern Modeling. *IEEE Computer Society*, 214-219.

[16] Valuer, F., Vigna, G., Kruegel, C., & A. Kemerrer, R. (2004). A Comprehensive Approach to Intrusion Detection Alert Correlation *IEEE Transaction on dependable and Secure Computing, 1*(3).

[17] Lincoln Lab, M. (1999). 1999 DARPA Intrusion Detection Evaluation Plan [Electronic Version].

[18] McAfee. (2003). Virus Profile: W32/Lovsan.worm.a [Electronic Version]. Retrieved 23 July 2009 from http://home.mcafee.com/VirusInfo/VirusProfile.aspx?key=100547.

[19] Symantec. (2003). W32.Blaster.Worm [Electronic Version]. Retrieved 23 July 2009 from http://www.symantec.com/security_response/writeup.jsp?docid=2003-081113-0229-99.

[20] Sachs, M. H. (2003). SANS Top-20 Entry: W5 Windows Remote Access Services [Electronic Version]. Retrieved 10 December 2009.

[21] Dübendorfer, T., Arno Wagner, Theus Hossmann, & Plattner, B. (2005, July 7-8). *Flow-Level Traffic Analysis of the Blaster and SobigWorm Outbreaks in an Internet Backbone.* Paper presented at the Detection of Intrusions and Malware & Vulnerability Assessment, IEEE, Vienna, Austria.

[22] Cliff, C. Z., Don, T., & Weibo, G. (2006). On the Performance of Internet Worm Scanning Strategies. *ACM Performance and Evaluation Journal, 63*(7), 700-723.

[23] Braverman, M. (2005). Win32/Blaster: A Case Study from Microsoft's Perspective. On *VIRUS BULLETIN CONFERENCE OCTOBER 2005*.

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

Avoiding Black hole and Cooperative Black hole Attacks in Wireless Ad hoc Networks

Abderrahmane Baadache
Laboratory of Industrial Technology and Information,
University of A. Mira, Targua Ouzemour, 06000,
Bejaia, Algeria.

Ali Belmehdi
Laboratory of Industrial Technology and Information,
University of A. Mira, Targua Ouzemour, 06000,
Bejaia, Algeria.

***Abstract*— In wireless ad hoc networks, the absence of any control on packets forwarding, make these networks vulnerable by various deny of service attacks (DoS). A node, in wireless ad hoc network, counts always on intermediate nodes to send these packets to a given destination node. An intermediate node, which takes part in packets forwarding, may behave maliciously and drop packets which goes through it, instead of forwarding them to the following node. Such behavior is called black hole attack. In this paper, after having specified the black hole attack, a secure mechanism, which consists in checking the good forwarding of packets by an intermediate node, was proposed. The proposed solution avoids the black hole and the cooperative black hole attacks. Evaluation metrics were considered in simulation to show the effectiveness of the suggested solution.**

Keywords- wireless ad hoc network, routing protocol, security, black hole, cooperative black hole.

I. INTRODUCTION

A wireless ad hoc network is a collection of nodes connected by wireless links. Although it offers the advantage of being easy to deploy, the wireless ad hoc network paradigm, characterized by the absence of any control at the routing operation or data forwarding, introduced truths problems of security making thus less powerful the operation of such network.

Early conceived routing protocols suppose saint the environment in which the network is deployed [1]. However, that it is the contrary in practice, and it proves to be difficult the guarantee of the security within a network, where the medium of communication is open, a central authority of certification misses, what facilitates the interception, the modification or even the manufacture of packets so that then to inject them into the network, in order to disturb the correct operation or to make entirely non operational the network. Guarantee the security is not limited to ensure, individually, the following services: Participants authentification, data integrity and confidentiality, non repudiation, access control to the communication medium and anonymity. But also, how to put individual solutions together to produce a solution whose range will hold in account the beforehand listed considerations and requirements. The literature contains security solutions which are protocols whose originating objective was security, and other protocols conceived at the beginning without security like primary objective, and they were secured in improved versions because attacks of which they were victims. Obviously, security remains always an arising problem and the remedy is far from being found.

Various attacks against wireless ad hoc networks can be conducted [1]. They are qualified passive ones, if they are limited to the listening of the network traffic to take note, or active if the traffic is modified by the intruder. Security attacks can be internal when the malicious node belongs to the network, or external if not. Deny of service attacks are easy to carry out, and difficult to detect. Their principle is the violation and the non respect of the network protocol specification and their finality is the disturbance of the correct network operation. The no relaying of the traffic (of control or data) by an intermediate node constitutes a behavioral deviation, whose consequence is the violation of the objective for which the network is deployed. Such malicious behavior is called the black hole attack. In this paper and after having specified how a black hole attack is conducted, a solution consisting in checking the good forwarding of the traffic by an intermediate node, was proposed. The solution is based on the well-known principle which is the Merkle tree [2], [4].

The rest of the paper is organized as follows: Section 2 summarizes the related work, follow-up by a necessary background in section 3, then the black hole attack specification is presented in section 4 and detailed description of the suggested solution is the subject of section 5. Simulation and results are analyzed and discussed in section 6. The paper is achieved by a conclusion and our future work.

II. RELATED WORK

They did not have satisfied solutions to solve the black hole attack problem, what led researchers to be addressed to this attack to find remedies for it. In the literature, the black hole attack solutions can be solutions which are interested in the black hole attack acting in an individual manner or those which are interested in the cooperative black hole attack or general

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

security mechanisms being interested in others attacks in addition to the black hole attack.

In [4], Hongsong et al. proposes an intrusion detection model to combat the black hole attack in AODV [5] routing protocol. In this model, a security agent, established by a hardware thread in network processor uses parallel multithreading architecture, try to detect two cases of figure of attack. Those exploiting AODV control messages RREQ (Route REQuest) and RREP (Route REPly). The agent monitors the RREQ-RREP messages at real-time and if any detection rule is violated, the black hole attack is detected and the malicious node is isolated and recorded to a black list. This solution requires a special material for its implementation. It is dedicated to AODV protocol and it considers only control messages, however that black hole attack can target data messages. Considering always AODV, authors of [6] try to detecting abnormality occurs during the black hole attack by defining a normal state from dynamic training data that is updated at regular time intervals. To express the state of the network, the following features are used: Number of sent out RREQ messages, Number of received RREP messages and the change of the sequence number value used by AODV to determine the route freshness degree. Through the simulation, this method shows significant effectiveness however a more processing overhead is needed for its implementation and consequently it can suffer from scalability problems. In [7], the authors are always interested in AODV and propose a solution in which the receiving node of RREP message compares the sequence number value with a dynamic updated threshold. If the sequence number value is found to be higher than the threshold value, the node is suspected to be malicious and it adds the node to the black list. Here still, except message RREP is controlled. It is necessary to hold in account also data packets because a black hole node can behave normally in the route establishment phase and maliciously in the data transmission phase. In more, the threshold considered can miss exactitude what brings back to false alarms.

A cooperative black hole attack is when several malicious nodes work together as a group. To identify multiple black hole nodes acting in cooperation, Fu's team, in [8], [9], proposes slightly modified AODV protocol by introducing Data Routing Information (DRI) table, that contains information on routing data packet from/through the node and cross checking process that determines the reliable nodes to discover secure paths from source to destination. In [10], authors propose an enhancement of the basic AODV routing protocol to combat the cooperative black hole attack. They use a structure, which they call fidelity table, wherein every participating node will be assigned a fidelity level that acts as a measure of reliability of that node. In case the level of any node drops to 0, it is considered to be a black hole node and is eliminated. In their approach, they assume that nodes are already authenticated which is a little strong assumption. Agrawal et al. [11] proposes a complete protocol to detect a chain of cooperating malicious nodes in an ad hoc network. The proposed protocol is based on sending equal and small sized blocks of data, and monitoring the traffic flow at the neighborhoods of both source and destination, then gathering results of monitoring by a trusted backbone network, with the little strong assumption that a neighborhood of any node has more trusted than malicious nodes.

Other black hole attack solutions were summarized in [1] and dedicated solutions for securing particular routing protocols against all possible attacks, not only against the black hole, were proposed. SAR [12] and SEAD [13] are examples of such secure routing protocols. The watchdog and pathrater [14] mechanism is also proposed for mitigating misbehavior. The watchdog for identifying misbehaving nodes and the pathrater helps routing protocols avoid these nodes. This mechanism is functional only with nodes equipped by interfaces that support promiscuous mode operation.

III. BACKGROUND

Our solution is based on the principle of Merkle tree [2], [3] and uses the AODV [5] routing protocol like case of study, thus the knowledge of Merkle tree principle and AODV functioning is necessary.

A. *Merkle tree*

A Merkle tree is a binary tree in which, each leaf carries a given value and the value of an interior node (including the root) is a one-way hash function of the node's children values. Figure 1 illustrates an example of a Merkle tree in which:

- *h* denotes a one-way hash function. For example, the function SHA-1 [15].
- || is the concatenation operator.
- Values of leaves 1,2,4, respectively, are: *h(a)*, *h(b)*, *h(c)*.
- The value of the interior node 3 is: *h(h(a)||h(b))* which is the hashing result of the concatenation of values of children 1 and 2. Idem for the node 5 whose value is *h(h(h(a)||h(b))||h(c))* and children are: 3 and 4.

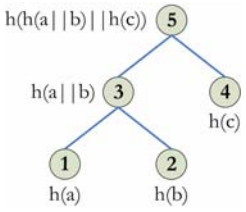


Figure 1. Example of a Merkle tree

B. AODV routing *protocol*

AODV (Ad hoc On-Demand Distance Vector) [5] is a reactive routing protocol composed of two modules:

- Route discovery module: To send data to a given destination *D*, the source node *S* consults its routing table. If it finds a valid entry (a route) towards this destination *D*, it uses it immediately, else it launches a route discovery procedure (see Figure 2.), witch consists in broadcasting, by the source node *S*, a route request (RREQ) message (containing amongst other information: destination's address, destination's sequence number) towards neighbors. When RREQ is

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

received by an intermediate node, this last consults its routing table to find a fresh route (the route is fresh if the sequence number of this route is larger than that of RREQ) towards the requested destination in RREQ. If such a route is found, a route reply (RREP) message is sent through the pre-established reverse route (established when RREQ pass through intermediate nodes) towards the source *S*. If the intermediate node does not find a fresh route, it updated its routing table and sends RREQ to these neighbors. This process is reiterated until RREQ reaches the destination node *D*. The destination node *D* sends RREP to *S* by using the pre-established reverse route. It should be noted that the source S can receive several RREP, it will choose that whose destination's sequence number is larger, if destination's sequence numbers of several RREP are equal, that of which the smallest hope counter will be selected.

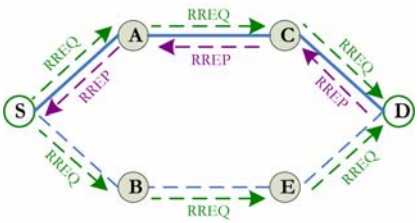


Figure 2. Route discovery process of AODV

- Route maintenance module: AODV uses Hello messages to maintain the connectivity between nodes. Each node periodically sends a Hello message to these neighbors and awaits Hello messages on behalf of these neighbors. If Hello messages are exchanged in the two directions, a symmetrical link between nodes is always maintained if no link interrupt occurs. The broken link can be repaired locally by the node upstream, else a route error (RERR) message is sent to the source *S* (see Figure 3.). This last can launch again, if necessary, the route discovery procedure. It should be noted that the link interrupt is the consequence of the mobility or the breakdown of nodes.

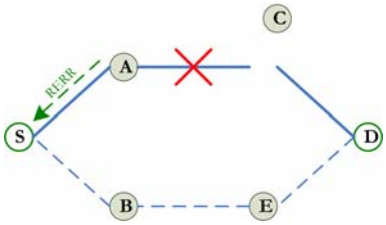


Figure 3. Route maintenance process of AODV

IV. BLACK HOLE ATTACK SPECIFICATION

In a black hole attack, the malicious node refuses to forward data packets to the following node in the route connecting a given source and destination. To conduct its attack, the malicious node must initially belong to the data route, then it pass to the action which is the data dropping. According to the specification of the target routing protocol, the manner with which the malicious node fits in the data route differs. Since our case of study is the AODV protocol, we will see how a malicious node can make a success of its attack in AODV.

Two kinds of black hole attack can be distinguished:

- Internal black hole attack: The malicious node is an internal node which does not seek to fit in an active route between a given source and destination, and if the chance would have it, this malicious becomes element of an active data route, it will be able to conduct its attack as the transmission of the data starts. This attack is internal because the malicious node belongs already to the data route. Here, there is no violation of AODV specification and the malicious node has anything to make to carry out with success its attack.

- External black hole attack: The malicious node is an external node which seeks to fit in an active route. For that, it violates the routing protocol specification and executes the process schematized in Figure 4. and summarized in the following points:
  - The malicious node detects the existence of an active route and takes note of the destination address.
  - The malicious node prepares a route replay packet (RREP) in which: the destination address field is set to the spoofed destination address, the sequence number is set to a greatest value and the counter hope is set to a smallest value.
  - The malicious node sends this route reply RREP to the nearest intermediate node belonging to the real active route (not necessarily to the data source node himself).
  - The route reply RREP received by the intermediate node will be relayed through the preestablished inverse route towards the data source node.
  - The source node updates its routing table by the new information received in the route reply.
  - The source uses the new route to sending data.
  - The malicious node starts to drop the data in the route to which it belongs.

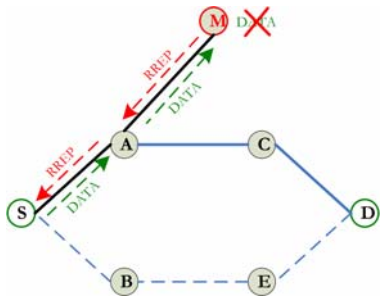


Figure 4. Black hole attack specification

The malicious node *M* detects the existence of active route (*S,A,E,D*) between the source *S* and destination *D*. The malicious node *M* sends to the node *A* a route reply message (RREP) containing the spoofed destination address, a relatively large sequence number and a small hope count. The node *A*

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

forwards this route reply message to *S*. This last updates its routing table and considers the new route (*S,A,M*). The node *S* uses this route to send data and while arriving at *M*, these data will be quite simply dropped. Nodes source and destination will not be able any more to communicate in the presence of the black hole attack.

V. OUR SOLUTION

The table I contains the notations used to describe our solution.

TABLE I. NOTATIONS

Notation	Significance
id_i	Identity of node i.
S_i	Secret generated by node i.
h	One-way hash function.
||	Concatenation operator.

In Figure 5, we consider a piece of network made up of 3 nodes *A*, *B* and *C*. On this last, a Merkle tree is juxtaposed. We point out that our goal is to check that *B* conveys well, towards *C*, the traffic sent by *A*.

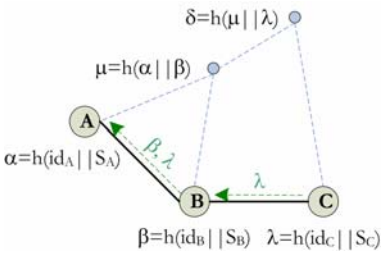


Figure 5. Basic principle of the solution, single black hole case

Each node *i* holds the value $h(id_i||S_i)$, $i \in \{A,B,C\}$. The transmitter node *A* has, in more, the value δ (value of the root of the Merkle tree). So that *A* checks that *B* forwards well the traffic to *C*, the node *C* sent the value λ (value held by *C*) to *B* and *B*, in turn, sent to *A*, β (value held by *B*) and λ.When both β and λ are received by *A*, the node *A* recalculates δ from α (value owned by A), β and λ, then compares the result with the value of δ already held, if equality, the node *B* routed out the traffic to *C*, otherwise, *B* is a black hole node. Obviously, to leading a black hole attack, the node *B* must generate the value λ which is impossible, because it does not know the *C*'s secret S_c.

Nodes *B* and *C* can cooperate to conduct black hole attack, this is easy if *C* communicates to *B* its secret S_c. To prevent such a scenario, the idea of the solution can be generalized as it is shown in Figure 6.

When β, λ and ω are received by *A*, the node *A* recalculates ψ from α, β, λ and ω, then compares the result with the value ψ of already held, if equality, the route (*A,B,C,D*) is secured, otherwise, the route contains a black hole nodes. *B* and *C* cannot conduct a cooperative black hole attack because they do not know the secret S_D held by *D*.

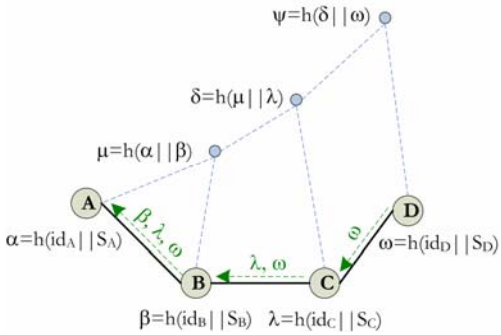


Figure 6. Basic principle of the solution, cooperative black hole case

It is important to note that the secret S_i is generated by node *i* independently of other nodes and this to prevent another node to replay the role of the node *i.*

Once detected, the black hole node can be put in a black list to avoid it in all future communication and, if necessary, another discovery route operation can be started again. For the implementation of our solution, an initialization process is necessary. It enables to nodes to generate each one its secret and communicate the root value of the Merkle tree to the transmitter node (communicate the value δ to the node *A* in Figure 5).

The gray hole attack is a variant of the black hole attack in which a malicious node, selectively, destroys packets of the traffic that passes through it. Our solution can easily be adapted to thwart such attacks. It is remarkable that this solution can be integrated in any routing protocol, at the time when its operation is not related to any specificity of a particular routing protocol. Obviously, the specification of the black hole attack is different from a routing protocol to another.

VI. SIMULATION AND ANALYSIS

In the network of the Figure 7, we use a source data node, a destination data node and one or more black hole nodes chosen randomly among nodes of the network. OPNET Modeler version 11.5 [16] is used as a simulator. The table II contains the simulation parameters.

TABLE II. SIMULATION PARAMETRES

Simulation parameter	Value
Nodes number	10
Network size	1km*1km
Simulation duration (sec)	600
Packet Inter-Arrival Time (sec)	exponential(1)
Packet size (bits)	exponential(1024)
Transmit Power (watt)	0.0001
Routing protocol	AODV
Hash function	SHA-1

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

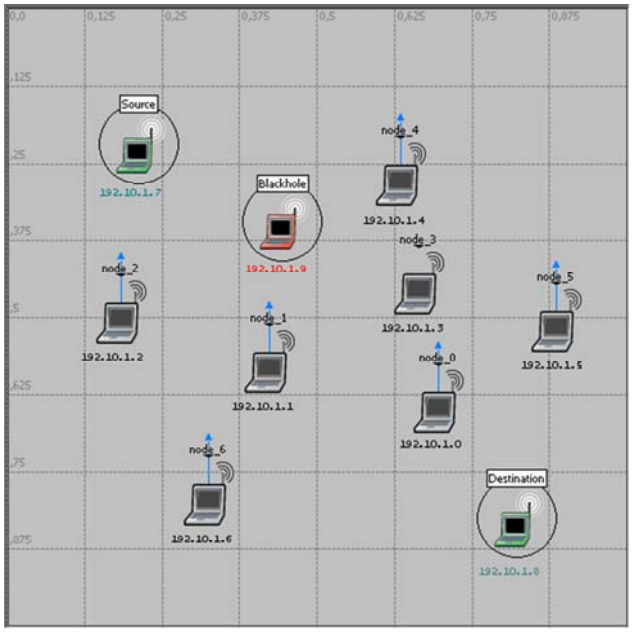


Figure 7. Network used in simulation

In this section, we show the effect of the black hole attack on the network functioning in first place, and how our solution can neutralize such an attack in second. This, in the presence of one or more black hole attacks (cooperative black hole). For this, we consider the following metrics:

- Traffic sent by the node source (packet/sec): indicate the number of packets/second sent by the source node.
- Traffic received by the destination node (packet/sec): indicate the number of packets/second received by the destination node.

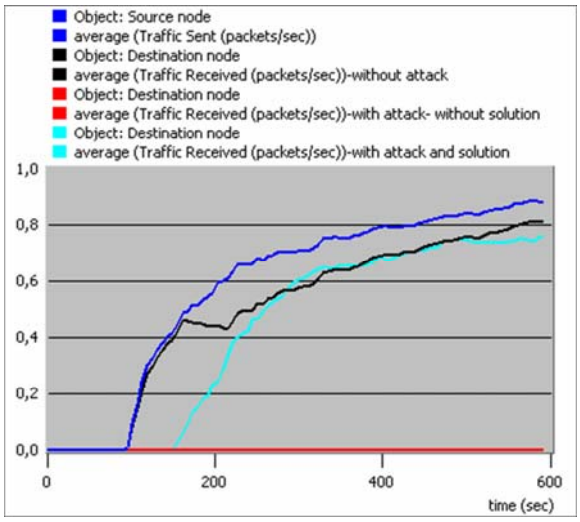


Figure 8. Traffic Sent & Traffic Received - Single Black hole

Figure. 8 shows the traffic sent by the source node, the traffic received by the destination node without attack, the traffic received by the destination node in the presence of a single black hole node and the traffic received by the destination node when the solution is used. The traffic received by the destination node is null under the effect of the attack without using the solution, this is justified by the destruction of packets per black hole node. Under the effect of the solution, le black hole node is eliminated and the destination node receives normally the traffic.

Figure 9. shows the simulation results in the presence of two malicious nodes cooperating together to conduct a black hole attack (cooperative black hole).

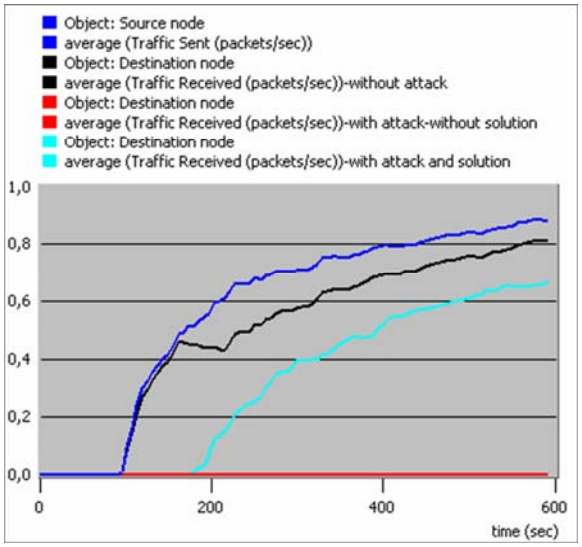


Figure 9. Traffic Sent & Traffic Received - Cooperative Black hole

Our solution detects well the cooperative black hole attack but with a little time lag because the detection of the cooperative black hole attack requires more calculation than in the case of a simple black hole, then an alternative route is found to convey the traffic to the destination node.

To evaluate the performance of our solution, we consider the following evaluation metrics:

- End to end delay (sec): indicate the delay, in second, for sending a bit of the source node to the destination node.
- Network load (bits/sec): indicate the traffic quantity, in bits/sec, in the entire network.

Figure 10. and Figure 11. show, respectively, the end to end delay and the network load, in cases : without black hole attack, with black hole attack and solution, and cooperative black hole attack and solution.

A very small time lag due to tests and calculations carried out by nodes, after which, the delay is stabilized and become almost identical in all scenarios, which shows that our solution does not influence the degradation of the network performance.

A slight increase in the network load because of messages exchanged between nodes to communicate various hash values. More messages in the case of cooperative black hole, comparatively with the case of single black hole, which justified graphs pace in Figure 11.

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

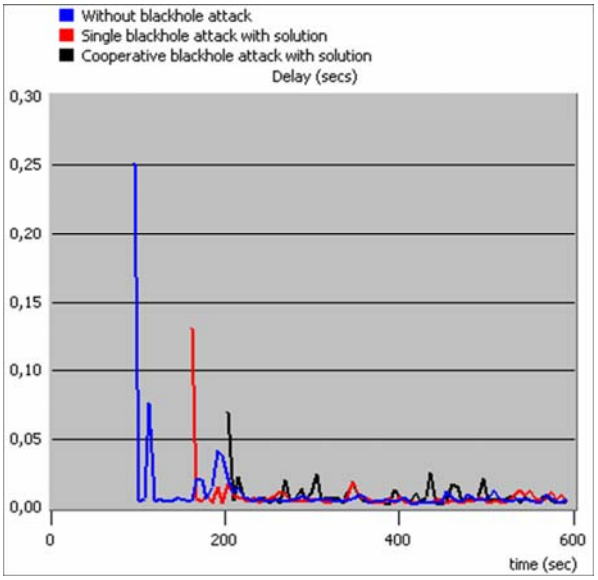


Figure 10. End to end delay

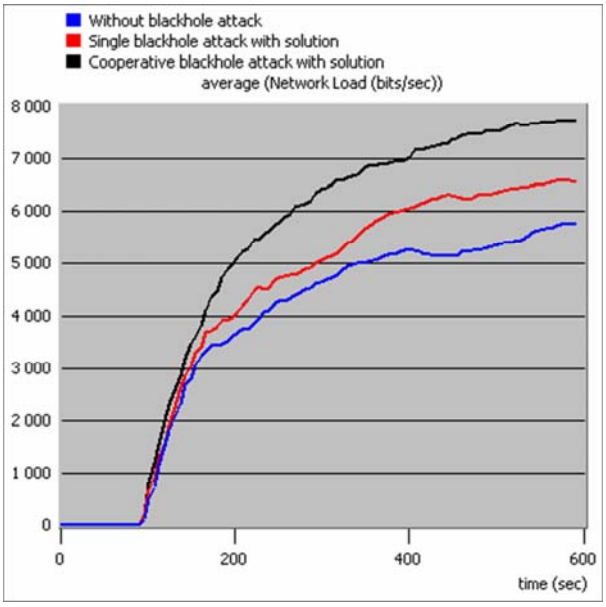


Figure 11. Network load

In the issue of these results, the network performance remains acceptable, if it is only a negligible variation in the end to end delay and a slight increase in the network load.

VII. CONCLUSION AND FUTUR WORK

Ad hoc network security is a serious problem by which researchers were concerned. Several security solutions are suggested but perfect security is far from beings obvious. We focused in this paper to the black hole attack, which refuses to convey the traffic and drop it. After the black hole attack specification in an example of routing protocol (AODV), we proposed a solution, based on the principle of Merkle tree, for avoiding the black hole and the cooperative black hole attacks. Moreover, Gray hole, a variant of the black hole, can easily be detectable by our solution. Simulation results show well its effectiveness in the detection of such attacks and the slight influence of this solution on the network performance. It was announced that our solution is general and it is not dedicated to any routing protocol.

We think that the network density, nodes mobility and the number of black hole nodes are determining factors in our solution performance, in term of end to end delay and network load. In a future work, we will study the influence of these factors and will find adequate mechanisms which will make the solution more powerful.

ACKNOWLEDGMENT

Colleagues, H. Moumen and S. Mustapha are thanked for the second reading of this paper and for pertinent criticisms and constructive remarks.

REFERENCES

[1] B. Kannhavong, H. Nakayama, Y. Nemoto, N. Kato, and A. Jamalipour, "A survey of routing attacks in mobile ad hoc networks," in *IEEE Wireless Communications*, Oct. 2007, pp. 85–91.

[2] A. J. Menezes, P. V. Oorschot, and S. A. Vanstone, "*Handbook of Applied Cryptography"*, CRC Press, 1996.

[3] J. L. Munoz, J. Forne, O. Espazara, and M. Soriano, "Certificate revocation system implementation based on the merkle hash tree," *International Journal of Information Security*, vol. 2, no. 2, pp. 110–124, Jan. 2004.

[4] C. Hongsong, J. Zhenzhou, and H. Mingzeng, "A novel security agent scheme for aodv routing protocol based on thread state transition," *Asian Journal of Information Technology*, vol. 5, no. 1, pp. 54–60, 2006.

[5] C. Perkins, E. B. Royer, and S. Das, "Ad hoc on-demand distance vector (aodv) routing," *RFC: 3561, Nokia Research Center*, 2003.

[6] S. Kurosawa, H. Nakayama, N. Kato, A. Jamalipour, and Y. Nemoto, "Detecting blackhole attack on aodv-based mobile ad hoc networks by dynamic learning method," *International Journal of Network Security*, vol. 5, no. 3, pp. 338–346, Nov. 2007.

[7] P. N. Raj and P. B. Swadas, "DPRAODV: A dyanamic learning system against blackhole attack in aodv based manet," *IJCSI International Journal of Computer Science Issues*, vol. 2, pp. 54–59, 2009.

[8] H. Weerasinghe and H. Fu, "Preventing cooperative black hole attacks in mobile ad hoc networks: Simulation implementation and evaluation," *International Journal of Software Engineering and Its Applications*, vol. 2, no. 3, pp. 39–54, Jul. 2008.

[9] S. Ramaswamy, H. Fu, M. Sreekantaradhya, J. Dixon, and K. Nygard, "Prevention of cooperative black hole attack in wireless ad hoc networks," in *International Conference on Wireless Networks (ICWN'03)*, Las Vegas, Nevada, USA, 2003.

[10] L. Tamilselvan and V. Sankaranarayanan, "Prevention of co-operative black hole attack in manet," *Journal of Networks*, vol. 3, no. 5, pp. 13–20, May 2008.

[11] P. Agrawal, R. K. Ghosh, and S. K. Das, "Cooperative black and gray hole attacks in mobile ad hoc networks," in *Proceedings of the 2nd International Conference on Ubiquitous Information Management and Communication (ICUIMC'08)*. New York, NY, USA: ACM, 2008, pp. 310–314.

[12] S. Yi, P. Naldurg, and R. Kravets, "Security-aware ad hoc routing for wireless networks," in Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing (MobiHoc'01). New York, NY, USA: ACM, 2001, pp. 299–302.

[13] Y. Hu, B. J. David, and P. Adrian, "Sead: secure efficient distance vector routing for mobile wireless ad hoc networks," *Journal of Ad Hoc Networks*, vol. 1, no. 1, pp. 175–192, Jul. 2003.

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

[14] S. Marti, T. J. Giuli, K. Lai, and M. Baker, "Mitigating routing misbehavior in mobile ad hoc networks," in *Proceedings of the 6th annual international conference on Mobile Computing and networking (MobiCom'00)*. New York, NY, USA: ACM, 2000, pp. 255–265.

[15] D. Eastlake and P. Jones, "Us secure hash algorithm 1 (sha1)," *RFC: 3174, Network Working Group*, Sep. 2001.

[16] Opnet website. [Online]. Available: http://www.opnet.com/

AUTHORS PROFILE

Abderrahmane Baadache obtained the *engineer* degree in computer science, option computing systems, from the national institute of computing (INI), Oued Smar, Algiers, Algeria in 2001 and the *magister* degree, option networks and distributed systems from the doctoral school of networking and distributed systems, department of computer science, A. Mira University, Béjaia, Algeria in 2005. Currently, he is preparing his PhD under the direction of the professor A. Belmehdi. His research topic is wireless networking, in particular, security in ad hoc, sensor, mesh and vehicular networks.

Ali Belmehdi is a professor at the University of A. Mira, Bejaia, Algeria. Besides his teaching activities, he is the header of the Laboratory of Industrial Technology and Information. His career is rich in scientific publications and supervising students in PhD or magister degree. His research topic includes the automatic and the computer science, particularly, wireless networking.

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

Design of Current Controller for Two Quadrant DC Motor Drive by Using Model Order Reduction Technique

K.Ramesh
EEE Department,
Velalar College of Engg. & Tech.,
Erode - 638012, India

K.Ayyar
EEE Department,
Velalar College of Engg. & Tech.,
Erode - 638012, India

Dr.A.Nirmalkumar
EEE Department,
BIT, Sathyamangalam,
India

Dr.G.Gurusamy
EEE Department,
BIT, Sathyamangalam
India

Abstract -In this paper, design of current controller for a two quadrant DC motor drive was proposed with the help of model order reduction technique. The calculation of current controller gain with some approximations in the conventional design process is replaced by proposed model order reduction method. The model order reduction technique proposed in this paper gives the better controller gain value for the DC motor drive. The proposed model order reduction method is a mixed method, where the numerator polynomial of reduced order model is obtained by using stability equation method and the denominator polynomial is obtained by using some approximation technique preceded in this paper. The designed controller's responses were simulated with the help of MATLAB to show the validity of the proposed method.

Keywords- Current controller, Model order reduction, Integral Square Error.

I. INTRODUCTION

DC motors used in many applications such as steel rolling mills, electric trains and robotic manipulators require current and speed controllers to perform tasks. Major problems in applying a conventional control algorithm in a controller design are the effects of nonlinearity in a DC motor. The nonlinear characteristics such as friction and saturation could degrade the performance of conventional controllers. Many advanced model-based control methods such as variable structure control and model reference adaptive have been developed to reduce these effects. However, the performance of these methods depends on the accuracy of system models and parameters. In this paper current controller of two quadrant DC motor drive is considered. The linear operation of DC motor drive was taken in to account in the design stage of current controller. In conventional design methods, some of simplification processes are considered to design the controller parameter values but where as in this proposed method, the model order reduction technique was introduced for the controller parameter design values.

Both in systems and control engineering and in numerical analysis, a wealth of model order reduction techniques have been developed. Balanced truncation, Krylov subspace methods, proper orthogonal decomposition and other SVD-based methods are just a few classes of methods that have been developed.

The computation of equivalent linear system models of large linear dynamic systems is a topic of considerable practical interest. This interest is motivated by the reduced complexity obtained by reducing the large linear sub-network in a linear (or nonlinear) network. Ideally, linear analysis on these sub-networks is performed by first computing a state space model or equivalent transfer function form, followed by the application suitable analysis method. However, the applicability of this method is limited since typical dynamic systems are represented by very large scale matrices that require specialized large-scale eigen analysis programs and computer resources. To avoid this practical limitation, model-order reduction methods are widely used in the solution of such systems. The basic idea behind model-order reduction is to replace the original system equations with a much smaller state-space or transfer function dimension. In particular, the identified reduced order model frequency characteristics must approximate those of the full order model.

In the analysis of many systems for which the physical laws are well known, one is frequently confronted with problems arising from the high dimensions of descriptive state model, the famous curse of dimensionality. The reduction of such high order systems (also termed as large scale systems) into low order models is one of the important problems in control and system theory system and is considered important in analysis, synthesis and simulation of practical systems. The exact analysis of high order systems is both tedious and costly.

To overcome the stability problem Hutton & Friedland [1] and Appiah [2] gave different methods, called stability based reduction methods which make use of some stability criterion. Other approaches in this direction include the methods such as Shamash [3] and Gutman, Mannerfelt & Molandor [4] which do not make use of any stability criterion but always lead to the stable reduced order models for stable systems. Bosley and Lees [5] and others have proposed a method of reduction based on the fitting of the time moments of the system and its reduced model but these methods have a serious disadvantage that the reduced order model may be unstable even though the original high order system is stable. Some combined methods are also given for example Shamash

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

[6], Chen, Chang and Han [7] and Wan [8] in which the denominator of the reduced order model is derived by some stability criterion method while the numerator of the reduced model is obtained by some other methods. [9].

In this paper, a new model order reduction method is proposed and its helps in finding the current controller gain value. Simulation results were shows the validity of the proposed method. The proposed model order reduction method is a mixed method, where the numerator polynomial of reduced order model is obtained by using the stability equation method and numerator polynomial is obtained by the method proposed in the paper [10].

II.DC MOTOR DRIVE

The control schematic of a two-quadrant converter-controlled separately-excited DC motor drive is shown in figure 1. The motor drive shown is a speed controller system. The thyristor bridge converter gets its ac supply through a three phase transformer and fast acting ac contactors. The dc output is fed to the armature of the dc motor. The field is separately excited, and the filed supply can be kept constant or regulated, depending on the need for the field weakening mode of operation. The DC motor has a tachogenerator whose output is utilized for closing the speed loop. The motor is driving a load considered to be frictional for this treatment. The output of the tachogenerator is filtered to remove the ripples to provide the signal, ω_{mr}. The speed command ω_r^* is compared to the speed signal to produce a speed error signal. This signal is processed through a proportional-plus-integral (PI) controller to determine the torque command. The torque command is limited, to keep within the safe current limits and the current command is obtained by proper scaling. The armature current command i_a^* is compared to the actual armature command i_a to have a zero current error. The PI controller produces the equivalent control signal V_c when an error signal is occurred. The control signal accordingly modifies the triggering angle α to be sent to the converter for implementation.

The current control loop of DC motor drive is shown in the figure 2. The DC machine contains an inner loop due to the induced emf. It is not physically seen; it is magnetically coupled. The inner current loop will cross this back-emf loop, creating a complexity in the development of the model. The inner current loop assures a fast current response and also limits the current to a safer level. The inner current loop makes the converter a linear current amplifier. The outer speed loop ensures that the actual speed is always equal to the commanded speed and that any transient is overcome with in the shortest feasible time without exceeding the motor and converter capability.

The operation of the closed loop speed controlled drive is explained from one or two particular instances of speed command. A speed from zero to rated value is recommended, and the motor is assumed to be at standstill. This will generate a large speed error and a torque command and in turn an armature current command. The armature current error will generate the triggering angle to supply a a preset maximum DC voltage across the motor terminals. The inner current loop will maintains the level permitted by its commanded value, producing a corresponding torque. As motor starts running, the torque and current are maintained at their maximum level, thus accelerating the motor rapidly.

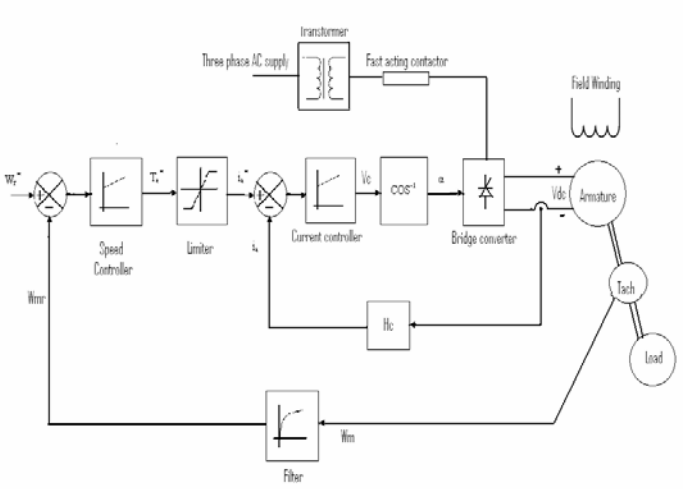


Fig.1. Speed-controlled two-quadrant DC motor drive

When the motor attains the commanded value, the torque command will settle down to a value equal to the sum of the load torque and other losses to keep the motor in steady state. The DC machine contains an inner loop due to the induced emf. It is not physically seen; it is magnetically coupled. The inner current loop cross this back-emf loop, creating a complexity in the development of the model and is shown in the fig.2.

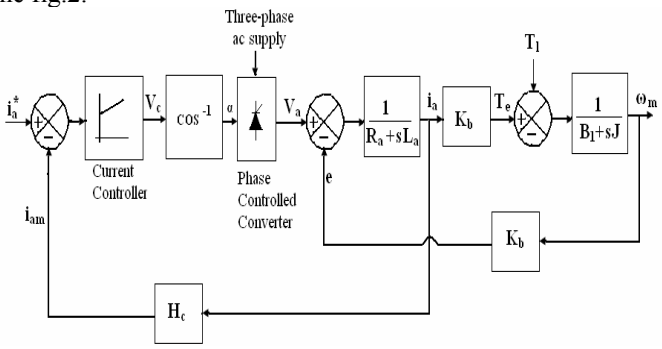


Fig.2. DC motor and current-control loop

The design of the gain constants of the current controllers is of paramount importance in meeting the dynamic specifications of the motor drive.

III.DESIGN OF CONTROLLERS

The overall closed-loop system of DC motor drive is shown in fig.3. The design of control loops starts from the innermost (fastest) loop to the outer (slowest) loop. The reason

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

to proceed from the inner to the outer loop in the design process is that the gain and time constants of only one controller at a time are solved, instead of solving for the gain and time constants of all controllers simultaneously.

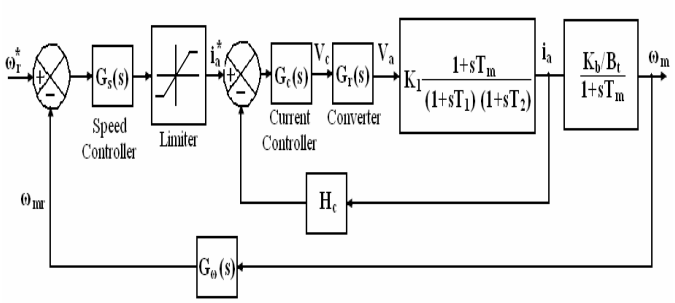


Fig.3. Block diagram of the DC motor drive

In some applications such as traction application, the motor drive need not be speed controlled but may be torque controlled. In that case, the current loop is essential and exists regardless of whether the speed control loop is going to be closed. The performance of the outer loop is dependent on the inner loop; therefore, the tuning of the inner loop has to precede the design and tuning of the outer loop.

IV.CURRENT CONTROLLER

The current control loop is shown in figure 4.

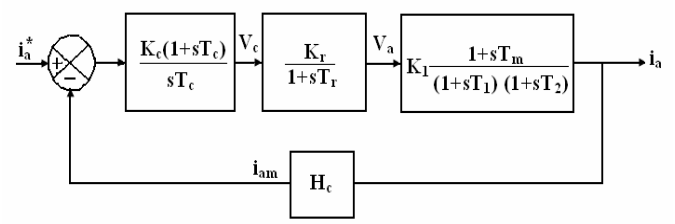


Fig.4. Current-control loop

The loop gain transfer function is,

$$GH_i(s) = \{\frac{K_1 K_C K_r H_C}{T_C}\}.\frac{(1+sT_c)(1+sT_m)}{s(1+sT_1)(1+sT_2)(1+sT_r)} \quad (1)$$

Where,

T_1, T_2= Electrical time constant of the motor,
B_t = Total friction co-efficient in N-M/ (rad/sec)
J = Moment of inertia, kg-m2
R_a = DC machine armature resistance
L_a = DC machine armature inductance
K_c = Current controller gain
T_c = Current controller time constant
K_r = Converter gain
T_r = Converter delay time
ω_m = DC motor speed.

This is a fourth order system, and simplification is necessary to synthesize a controller without resorting to a computer. Noting that T_m is on the order of a second and in the vicinity of the gain crossover frequency, we see that the following approximation is valid.

$$(1+sT_m) \cong sT_m \quad (2)$$

This reduces the loop gain function to

$$GH_i(s) = \frac{K(1+sT_c)}{(1+sT_1)(1+sT_2)(1+sT_r)} \quad (3)$$

Where,

$$K = \frac{K_1 K_C K_r H_C T_m}{T_C} \quad (4)$$

The time constants in the denominator are seen to have the relationship

$$T_r < T_2 < T_1 \quad (5)$$

The equation (3) can be reduced to second order, to facilitate a simple controller synthesis, by judiciously selecting

$$T_c = T_2 \quad (6)$$

Then the loop function is

$$GH_i(s) \cong \frac{K}{(1+sT_1)(1+sT_r)} \quad (7)$$

The conventional design methods uses the approximations used in the equations from (2)-(7).We can't assure that the approximations were made are good in the design aspect. So the proposed Model order reduction technique is applied to get the equivalent reduced order model of order two. The validity of the proposed method is verified through the time and frequency domain analysis.

The characteristic equation or denominator of the transfer function between the armature current and its command is

$$(1+sT_1)(1+sT_r)+K=0 \quad (8)$$

This equation is expressed in standard form as

$$T_1 T_r\left(s^2 + s\left(\frac{T_1+T_r}{T_1 T_r}\right) + \frac{K+1}{T_1 T_r}\right) = 0 \quad (9)$$

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

The general second order characteristic equation is given by,

$$s^2 + 2\xi\omega_n s + \omega_n{}^2 = 0 \tag{10}$$

On comparing the equations (9) and (10) with the damping ratio ξ as 0.707 for satisfactory operation of system, we get the value for the term K. From that value, the current controller gain K_c can be easily estimated.

V.MODEL ORDER REDUCTION METHOD

Let the nth order system is given by the transfer function

$$G(s)=\frac{\sum_{i=0}^{n-1} d_i s^i}{\sum_{j=0}^{n} e_j s^j}=\frac{N(s)}{D(s)}=\frac{a_0+a_1 s+a_2 s^2+.....+a_m s^m}{b_0+b_1 s+b_2 s^2+.....+b_{n-1}s^{n-1}+b_n s^n} \tag{11}$$

and the corresponding r^{th} (r<n) order reduced order is of the form

$$\mathrm{Gr(s)}=\frac{\overline{N(s)}}{\overline{D(s)}}=\frac{d_0+d_1 s+d_2 s^2+.....+d_{r-1}s^{r-1}}{e_0+e_1 s+e_2 s^2+.....+e_{r-1}s^{r-1}+e_r s^r} \tag{12}$$

The proposed model order reduction method consists of three steps

Step 1: *The denominator polynomial of reduced order model is obtained by using the stability equation method [11].*

For stable G(s) the eve and odd parts of D(s) may be factored as the following stability equations.

$$D_e(s) = e_0 \prod_{i=1}^{n/2}\left(1+\frac{s^2}{z_i^2}\right) \tag{13}$$

$$D_o(s) = e_1 \prod_{i=1}^{(n-1)/2}\left(1+\frac{s^2}{p_i^2}\right) \tag{14}$$

Where,

$$z_1^2 < p_1^2 < z_2^2 < p_2^2 < .....$$

After discarding the factors with larger magnitude of zi and pi, the stability equations are reduced, which are combined to give a reduced polynomial of order r, as:

$$D_r(s)=b_0\prod_{i=1}^{r/2}\left(1+\frac{s^2}{z_i^2}\right)+b_1\prod_{i=1}^{(r-1)/2}\left(1+\frac{s^2}{p_i^2}\right)=\sum_{j=0}^{r} b_j s^j \tag{15}$$

Step 2: *The numerator polynomial of reduce order system is obtained by using the method proposed in* [10].

Consider the given system transfer function given in (11)

$$\mathrm{G(s)}=\frac{a_0+a_1 s+a_2 s^2+.....+a_m s^m}{b_0+b_1 s+b_2 s^2+.....+b_{n-1}s^{n-1}+b_n s^n}$$
$$= K\frac{(1+A_1 s+A_2 s^2+.....+A_m s^m)}{(1+B_1 s+B_2 s^2+.....+B_{n-1}s^{n-1}+B_n s^n)};\mathrm{n}\geq\mathrm{m} \tag{16}$$

Where, $K=\frac{a_0}{b_0}$,

$A_i = \frac{a_i}{a_0}$; i = 0,1,2,3…m and

$B_j = \frac{b_j}{b_0}$; j = 0,1,2,3…n

Let the transfer function of the approximating low-order system be,

$$\mathrm{Gr(s)} = K\frac{(1+C_1 s+C_2 s^2+.....+C_q s^q)}{(1+D_1 s+D_2 s^2+.....+D_{p-1}s^{p-1}+D_p s^p)} \tag{17}$$

where n≥p≥q

The coefficients of D_1, D_2 , .. were obtained from the equation (15).

The following relation should be satisfied as closely as possible

$$\frac{|G(j\omega)|^2}{|G_r(j\omega)|^2}=1 \text{ for } 0\leq\omega\leq\infty \tag{18}$$

$$\frac{\mathrm{G(s)}}{\mathrm{G_1(s)}}=\frac{(1+A_1 s+A_2 s^2+.....+A_m s^m)(1+D_1 s+D_2 s^2+.....+D_{p-1}s^{p-1}+D_p s^p)}{(1+B_1 s+B_2 s^2+.....+B_{n-1}s^{n-1}+B_n s^n)(1+C_1 s+C_2 s^2+.....+C_q s^q)}$$

$$= \frac{(1+m_1 s+m_2 s^2+.....+m_u s^u)}{(1+l_1 s+l_2 s^2+.....+l_{v-1}s^{v-1}+l_v s^v)} \tag{19}$$

Where, u=m+p and v=n+q

$$\frac{|G(j\omega)|^2}{|G_r(j\omega)|^2}=\left.\frac{G(s)G(-s)}{G_r(s)G_r(-s)}\right|_{s=j\omega} \tag{20}$$

The equation produces the even powers of‘s’ and can be written as,

$$\frac{|G(j\omega)|^2}{|G_r(j\omega)|^2}=\left.\frac{1+L_2 s^2+L_4 s^4+.....+L_{2u}s^{2u}}{1+M_2 s^2+M_4 s^4+.....+M_{2v}s^{2v}}\right|_{s=j\omega} \tag{21}$$

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

$$\frac{|G(j\omega)|^2}{|G_r(j\omega)|^2} = 1 + \left.\frac{(L_2 - M_2)s^2 + (L_4 - M_4)s^4 + \ldots\ldots + (L_{2u} - M_{2v})s^{2u}}{1 + M_2 s^2 + M_4 s^4 + \ldots\ldots + M_{2v} s^{2v}}\right|_{s=j\omega} \quad (22)$$

To satisfy the above equation,

$L_2 = M_2$

$L_4 = M_4$

.

.

.

$L_{2u} = M_{2v}$; if u=v (23)

If u<v, then the error generated by the lower order model is,

$$|\varepsilon| = \frac{|G(j\omega)|^2}{|G_r(j\omega)|^2} - 1 \quad (24)$$

From the above equations, we can obtain the conditions as

$$L_{2x} = \sum_{i=0}^{x-1} (-1)^i 2m_i m_{2x-i} + (-1)^x m_x^{\,2} \quad (25)$$

for x= 1,2, 3……..u and m_0=1

and

$$M_{2y} = \sum_{i=0}^{y-1} (-1)^i 2l_i l_{2y-i} + (-1)^x l_y^{\,2} \quad (26)$$

for y= 1,2, 3……..v and l_0=1

From the equation (11), the reduced order model numerator coefficients can be obtained. Finally the reduced order model is in the form of

$$\mathrm{Gr(s)} = K \frac{(1 + C_1 s + C_2 s^2 + \ldots\ldots + C_q s^q)}{(1 + D_1 s + D_2 s^2 + \ldots\ldots + D_{p-1} s^{p-1} + D_p s^p)} \quad (27)$$

Step 3: *The coefficients of the reduced order denominator polynomial is adjusted further to get the better approximation with original system. The coefficient of 's' term in the denominator is increased by n% and the same was reduced from the coefficient of the term 's^2'. The value of n is choosen by trial and error method. Normally the value of n is ranging between 1 to15.*

VI.DESIGN EXAMPLE

The motor parameters and ratings of a speed controlled DC motor drive maintaining the field flux constant are as follows 220V, 8.3A, 1470 rpm, R_a=4Ω, J=0.0607kg-m^2, L_a=0.072H, B_t=0.0869N-m/rad/sec, K_b=1.26V/rad/sec. The converter is supplied from 230V, 3-PhaseAC at 50Hz. The converter is linear, and its maximum control input voltage is ±10V. The tachogenerator has the transfer function $G_\omega(s) = \frac{0.065}{1 + 0.002s}$. The speed reference voltage has a maximum of 10V. The maximum current permitted in the motor is 20A.

Converter transfer function:

$$K_c = \frac{1.35V}{V_{cm}} = \frac{1.35 \times 230}{10} = 31.05$$

V_{dc}(max) = 310.5 V

The rated DC voltage required is 220V, which corresponds to a control voltage of 7.09V. The transfer function of the converter is,

$$G_r(s) = \frac{31.05}{(1 + 0.00138s)} \quad (28)$$

Current transducer gain: The maximum safe control voltage is 7.09V, and this has to correspond to the maximum current error:

i_{max} = 20A

$$H_c = \frac{7.09}{I_{\max}} = 0.355V/A$$

Motor transfer function:

$$K_1 = \frac{B_t}{K_b^2 + R_a B_t} = 0.0449$$

T1=0.1077 sec and T2=0.0208 sec

$$T_m = \frac{J}{B_t} = 0.7\,\mathrm{sec}$$

The subsystem transfer functions are

$$\frac{I_a(s)}{V_a(s)} = K_1 . \frac{(1 + sT_m)}{(1 + sT_1)(1 + sT_2)} = \frac{0.0449(1 + 0.7s)}{(1 + 0.0208s)(1 + 0.1077s)} \quad (29)$$

$$\frac{\omega_m(s)}{I_a(s)} = \frac{K_b / B_t}{(1 + sT_{m)}} = \frac{14.5}{(1 + 0.7s)} \quad (30)$$

Design of current controller:

T_2 = 0.0208 sec; Tc=0.03 sec

By applying the proposed model order reduction technique, with the value of ζ=0.707, the value of K is obtained as,

K= 357.192

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

$$K_C = \frac{KT_C}{K_1 H_C K_r T_m} = 35.719$$

By using trial and error method the value of Kc can be adjusted in to the value of 3.1 to obtain the better response for the current controller. The validity of this proposed method is evaluated by plotting the frequency response of the closed loop current to its command. This is shown in fig.5.

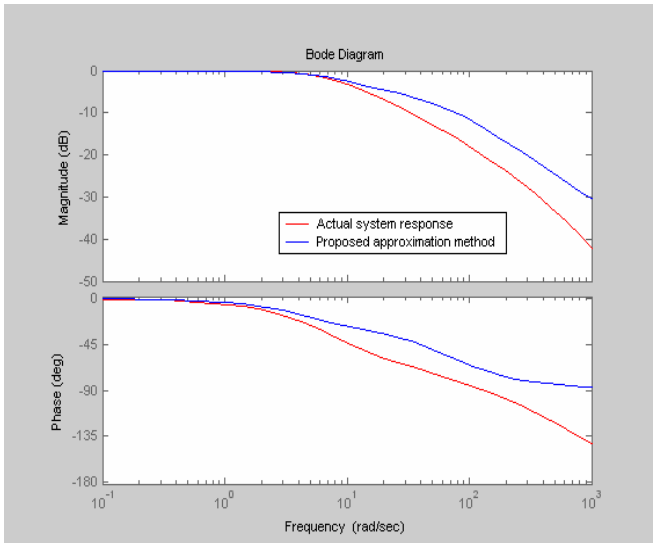


Fig.5 Frequency responses of actual and reduced order systems

From this figure, it is evident that the proposed method is quite valid in the frequency range of interest. The same in time domain analysis is shown in fig.6.

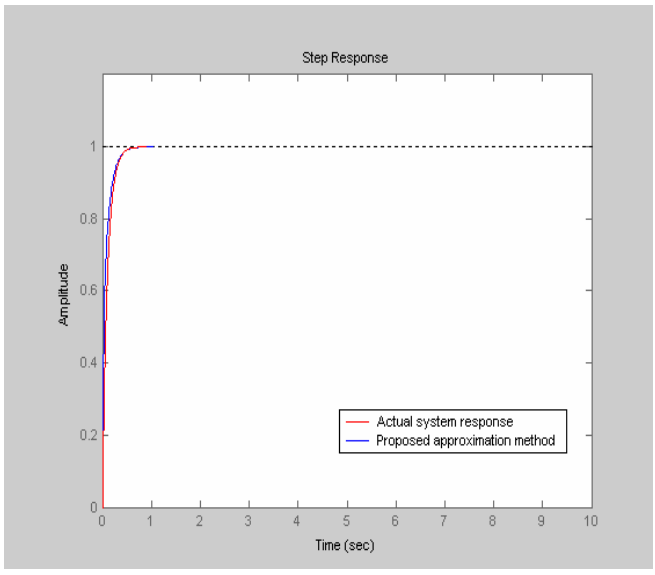
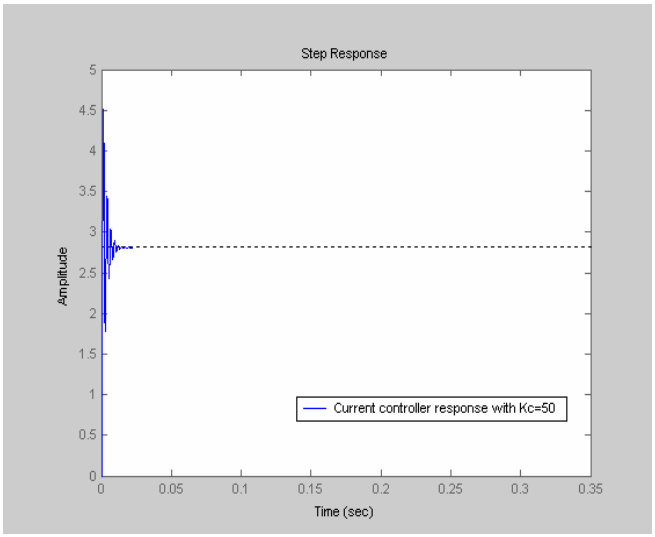
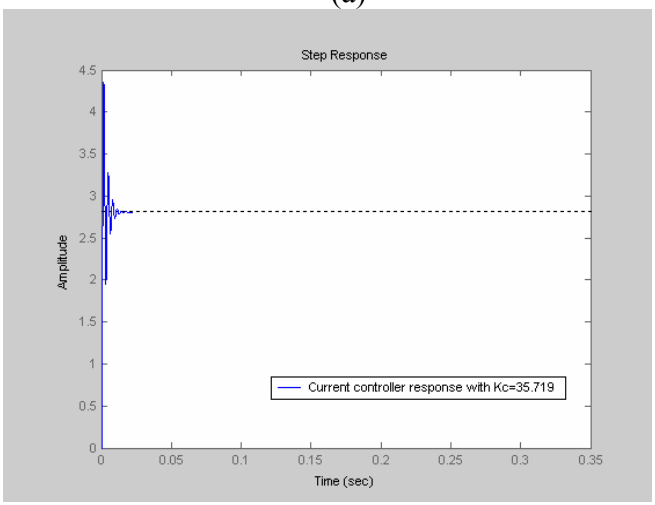


Fig.6 Time responses of actual and reduced order systems

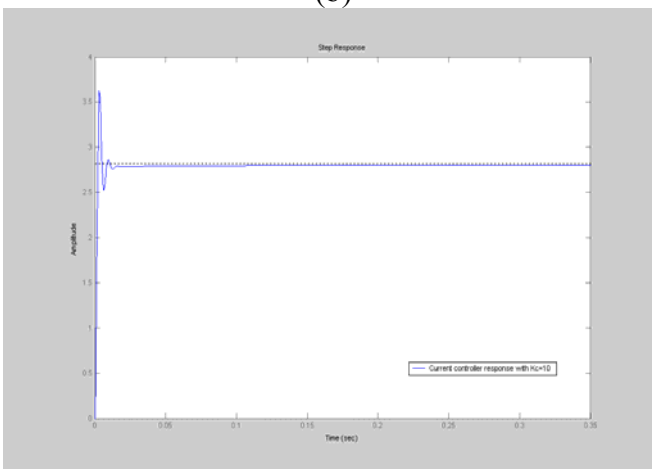
The time responses are important to verify the design of the controller and it is shown in fig. 7 for different values of current controller gain, K_c.



(a)

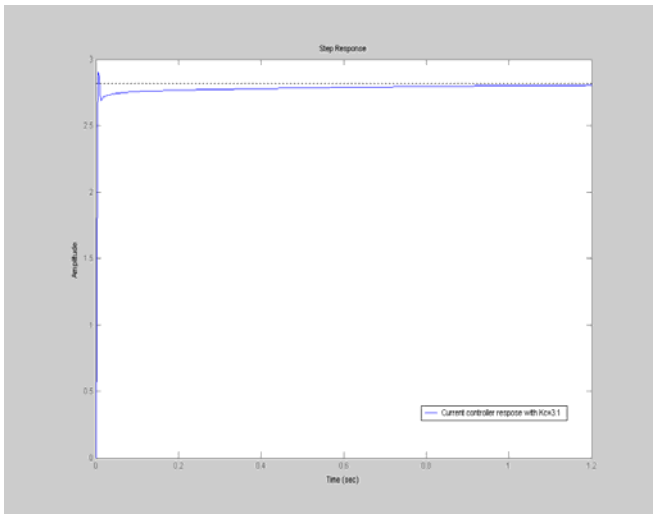


(b)



(c)

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



(d)

Fig.7 (a) Current controller response with K_c=50 (b) Current controller response with K_c=35.719 (c) Current controller response with K_c=10 (d) Current controller response with K_c=3.1

The value of K_c can be optimized by trail and error method. The integral square error (ISE) was calculated in between the actual and approximated system as 0.0204 and this shows the validity of the proposed method.

VII.CONCLUSION

The design of the gain constant of current controller is of paramount importance in meeting the dynamic specifications of the motor drive. This paper helps to obtain the same with the help of model order reduction technique. The validity of the proposed method was illustrated through the design example. The proposed model order reduction method is mathematically simple and produces the stable reduced order system if the given system is stable. The optimal value of current controller gain to suppress the oscillations at the output was obtained by trail and error method but it may be selected optimally by applying the neural network concepts or genetic algorithm.

REFERENCES

[1] M.F. Hutton and B. Friedland, "Routh approximations for reducing order linear time invariant systems", IEEE Trans. on Auto. Control, Vol.20, pp. 329-337, 1975.

[2] R.K. Appiah, "Linear model reduction using Hurwitz polynomial approximation", Int. J. Control, Vol.28, No.3, pp. 477-488, 1978.

[3] Y. Shamash, "Truncation method of reduction: a viable alternative", Electronics Letters, Vol.17, pp. 97-99, 1981.

[4] P.O. Gutman, C.F. Mannerfelt and P. Molander, "Contributions to the model reduction problem", IEEE Trans. on Auto. Control, Vol.27, pp.454-455, 1982.

[5] M.J. Bosley and F.P. Lees, "A survey of simple transfer-function derivations from high order state variable models", Automatica, Vol.8, pp. 765-775, 1972.

[6] Y. Shamash, "Model reduction using the Routh stability criterion and the Pade-approximation technique" Int. J. Control, Vol.21, pp. 475-484, 1975.

[7] T.C. Chen, C.Y. Chang and K.W. Han, "Reduction of transfer functions by the stability equation method", Journal of Franklin Institute, 308, pp. 389-404, 1979.

[8] Bai-Wu Wan, "Linear model reduction using Mihailov criterion and Pade approximation technique", Int. J. Control, Vol.33, pp. 1073-1089, 1981.

[9] T.C. Chen, C.Y. Chang and K. W. Han, "Model Reduction using the stability-equation method and the Pade approximation method", Journal of Franklin Institute, Vol.30, No.9, pp. 473-490, 1980.

[10] K. Ramesh, A. Nirmalkumar and G. Gurusamy, "Design of Digital IIR filters with the Advantages of Model Order Reduction Technique", International Journal of Electronics, Communications and Computer Engineering, Vol.1, No.2,pp 117-122, 2009

AUTHORS PROFILE

Ramesh.K received B.E. degree in Electrical and Electronics Engineering from K.S.R.College of Technology, Tamilnadu, India in 2002 and M.E. in Applied Electronics from Kongu Engg. College, Anna University, Chennai in 2005 and also received the MBA in System from PRIDE, Salem (Tamilnadu), India in 2005. Since then, he is working as a Assistant Professor in Velalar College of Engineering and Technology (Tamilnadu), India. Presently he is a Part time (external) Research Scholar in the Department of Information and Communication at Anna University, Chennai (India). His fields of interests include Model order reduction, Controller design and Optimization Techniques.

Ayyar.K received B.E. Degree in Electrical and Electronics Engineering from Ranipettai Engineering College, Tamilnadu, India in 2003 and M.E. in Power Electronics and Drives from Government College of Engineering, Salem, Anna University, Chennai in 2007. Since then, he is working as a Lecturer in Velalar College of Engineering and Technology (Tamilnadu), India. Presently he is a Part time (Internal) Research Scholar in the Department of Electrical and Electronics Engineering at Anna University, Coimbatore (India). His fields of interests include Power Drives and control, Controller design and System Optimization.

Dr.Nirmalkumar.A, received the B.Sc.(Engg.) degree from NSS College of Engineering, Palakkad in 1972, M.Sc.(Engg.) degree from Kerala University in 1975 and completed his Ph.D. degree from PSG Tech in 1992. Currently, he is working as a Professor and Head of the Department of Electrical and Electronics Engineering in Bannari Amman Institute of Technology, Sathyamangalam, Tamilnadu, India.

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

His fields of Interest are Power quality, Power drives and control and System optimization.

Dr.G.Gurusamy obtained his Pre-University education at St Johns College, Palayamkottai, and Thirunelveli district. He then joined PSG College of Technology, Coimbatore, in the year 1962 to pursue his Engineering course. He was graduated in Electrical Engineering in 1967. He latter obtained M.E (Applied Electronics) in 1972 and Ph.D in Control Systems in 1983. He is currently working as a Dean of Department of Electrical and Electronics Engineering in Bannari Amman Institute of Technology, Sathyamangalam. His fields of Interest are Advanced control, Digital control, Optimization and Biomedical electronics.

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

WIRELESS CONGESTION CONTROL PROTOCOL FOR MULTIHOP AD HOC NETWORKS

Mahendra kumar.S
Department of Electronics and Communication Engineering
Velalar College of Engineering and Technology
Tamil Nadu, India.

Senthil Prakash.K
Department of Electronics and Communication Engineering
Velalar College of Engineering and Technology
Tamil Nadu, India.

***Abstract*— The traditional TCP congestion control mechanism encounters a number of new problems and suffers a poor performance when the IEEE 802.11 MAC protocol is used in multihop ad hoc networks. Many of the problems result from medium contention at the MAC layer. In this paper, I first illustrate that severe medium contention and congestion are intimately coupled, and TCP's congestion control algorithm becomes too coarse in its granularity, causing throughput instability and excessively long delay. Further, we illustrate TCP's severe unfairness problem due to the medium contention and the tradeoff between aggregate throughput and fairness. Then, based on the novel use of channel busyness ratio, a more accurate metric to characterize the network utilization and congestion status,**

I propose a new wireless congestion control protocol (WCCP) to efficiently and fairly support the transport service in multihop ad hoc networks. In this protocol, each forwarding node along a traffic flow exercises the inter-node and intra-node fair resource allocation and determines the MAC layer feedback accordingly. The end-to-end feedback, which is ultimately determined by the bottleneck node along the flow, is carried back to the source to control its sending rate. Extensive simulations show that WCCP significantly outperforms traditional TCP in terms of channel utilization, delay, and fairness, and eliminates the starvation problem.

Keywords-component; Medium access control; Congestion control; Fairness; Multihop ad hoc networks; WCCP;

I. INTRODUCTION

Wireless ad hoc networks have found many applications in battlefield, disaster rescue and conventions, where fixed communications infrastructures are not available and quick network configurations are needed. To provide reliable transport services over and hence fully exploit the potential of ad hoc networks, efficient congestion control is of paramount importance. Unfortunately, the traditional TCP congestion control mechanism performs very poorly. As shown in recent studies [4].

TCP congestion control has an implicit assumption, i.e., any packet loss is due to network congestion. However, this assumption is no longer valid in the ad hoc networks as packet losses may well be due to channel errors, medium contention, and route failures.

In this paper, I mainly focus on the problems arising from medium contention when the IEEE 802.11 MAC protocol is used in the multi-hop ad hoc networks. I show that a rate based congestion control protocol is more appropriate than its window based counterpart in multihop ad hoc networks. We illustrate the close coupling between congestion and medium contention, which explains the instability of TCP. We also find that the optimum congestion window size of TCP may be less than one even in a very simple topology, say chain topology, in order to maximize the end-to-end throughput and minimize the end-to-end delay. Thus TCP tends to overshoot the network capacity and its granularity of sending rate adjustment is too coarse because the minimum increase in window size is the size of one packet upon each TCP acknowledgment or during each round trip time. Then we further show that medium contention also results in severe unfairness and starvation problems for TCP flows. Therefore, we conclude that congestion control, fairness and medium contention are all closely coupled in multihop ad hoc networks.

I propose a new wireless congestion control protocol (WCCP) based on the channel busyness ratio. In this protocol, each forwarding node determines the inter-node and intra-node fair channel resource allocation and allocates the resource to the passing flows by monitoring and possibly overwriting the feedback field of the data packets according to its measured channel busyness ratio. The feedback is then carried back to the source by the destination, which copies it from the data packet to its corresponding acknowledgment. Finally, the source adjusts the sending rate accordingly. Clearly, the sending rate of each flow is determined by the channel utilization status at the bottleneck node. In this way, WCCP is able to approach the max-min fairness in certain scenarios. We compare WCCP with TCP through extensive simulations in Section IV. We observe that WCCP significantly outperforms TCP in terms of channel utilization, delay, and fairness. Especially, it solves the starvation problem suffered by TCP.

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

A. MAC layer

The Media Access Control data communication protocol sub-layer, also known as the Medium Access Control, is a sublayer of the Data Link Layer. It provides addressing and channel access control mechanisms that make it possible for several terminals or network nodes to communicate within a multipoint network [3][5], typically a local area network or metropolitan area network . The hardware that implements the MAC is referred to as a Medium Access Controller. The MAC sub-layer acts as an interface between the Logical Link Control sub layer and the network's physical layer. The MAC layer emulates a full-duplex logical communication channel in a multipoint network. This channel may provide unicast, multicast or broadcast communication service.

B. AODV

Ad hoc On-Demand Distance Vector Routing is a routing protocol for mobile ad hoc networks and other wireless ad-hoc networks[2][4]. AODV is capable of both unicast and multicast routing. It is a reactive routing protocol, meaning that it establishes a route to a destination only on demand. In contrast, the most common routing protocols of the Internet are proactive, meaning they find routing paths independently of the usage of the paths. AODV is, as the name indicates, a distance-vector routing protocol. AODV avoids the counting-to-infinity problem of other distance-vector protocols by using sequence numbers on route updates, a technique pioneered by DSDV.

II. LITERATURE SURVEY

Good and bad network performance is largely dependent on the effective implementation of network protocols. TCP, easily the most widely used protocol in the transport layer on the Internet (e.g. HTTP, TELNET, and SMTP), plays an integral role in determining overall network performance. Amazingly, TCP has changed very little since its initial design in the early 1980's. A few "tweaks" and "knobs" have been added, but for the most part, the protocol has withstood the test of time. However, there are still a number of performance problems on the Internet and fine tuning TCP software continues to be an area of work for a number of people. The design of TCP was heavily influenced by the end-to end argument. Method of handling congestion and network overload. Are the key components of the end-to-end argument.

The four congestion control algorithms used in TCP are, Slow Start, Congestion Avoidance, Fast Retransmit and Fast Recovery. Slow Start, a requirement for TCP software implementations is a mechanism used by the sender to control the transmission rate [3], otherwise known as sender-based flow control. This is accomplished through the return rate of acknowledgements from the receiver. In other words, the rate of acknowledgements returned by the receiver determines the rate at which the sender can transmit data. When a TCP connection first begins, the Slow Start algorithm initializes a congestion window to one segment, which is the maximum segment size (MSS) initialized by the receiver during the connection establishment phase. When acknowledgements are returned by the receiver, the congestion window increases by one segment for each acknowledgement returned. Thus, the sender can transmit the minimum of the congestion window and the advertised window of the receiver, which is simply called the transmission window.

In the Congestion Avoidance [3] [9] algorithm a retransmission timer expiring or the reception of duplicate ACKs can implicitly signal the sender that a network congestion situation is occurring. The sender immediately sets its transmission window to one half of the current window size (the minimum of the congestion window and the receiver's advertised window size), but to at least two segments. If congestion was indicated by a timeout, the congestion window is reset to one segment, which automatically puts the sender into Slow Start mode. If congestion was indicated by duplicate ACKs, the Fast Retransmit and Fast Recovery algorithms [3] are invoked. As data is received during Congestion Avoidance, the congestion window is increased. However, Slow Start is only used up to the halfway point where congestion originally occurred. This halfway point was recorded earlier as the new transmission window. After this halfway point, the congestion window is increased by one segment for all segments in the transmission window that are acknowledged. This mechanism will force the sender to more slowly grow its transmission rate, as it will approach the point where congestion had previously been detected.

The Media Access Control data communication protocol sub-layer, also known as the Medium Access Control, is a sub layer of the data link layer[3]. It provides addressing and channel access control mechanisms that make it possible for several terminals or network nodes to communicate within a multipoint network, typically a local area network or metropolitan area network. The MAC sub-layer acts as an interface between the Logical Link Control sub layer and the network's physical layer. The MAC layer emulates a full-duplex logical communication channel in a multipoint network. This channel may provide unicast, multicast or broadcast communication service.

The channel access control mechanisms provided by the MAC layer are also known as a multiple access protocol. The most widespread multiple access protocol is the contention based CSMA/CD protocol used in Ethernet networks. This mechanism is only utilized within a network collision domain, for example an Ethernet bus network or a hub network. An Ethernet network may be divided into several collision domains, interconnected by bridges and switches [3][5]. A multiple access protocol is not required in a switched full-duplex network, such as today's switched Ethernet networks, but is often available in the equipment for compatibility reasons.

Several works have pointed out that greedy TCP can result in severe congestion in ad hoc networks and hence performance degradation. RED is one of the schemes proposed to overcome this problem [2]. Random early detection, also known as random early discard or random early drop is an active queue management algorithm. It is also a congestion avoidance algorithm [6].

In the traditional tail drop algorithm, a router or other network component buffers as many packets as it can, and

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

simply drops the ones it cannot buffer. If buffers are constantly full, the network is congested. Tail drop distributes buffer space unfairly among traffic flows. Tail drop can also lead to TCP global synchronization as all TCP connections "hold back" simultaneously, and then step forward simultaneously. Networks become under-utilized and flooded by turns. RED addresses these issues. It monitors the average queue size and drops (or marks when used in conjunction with ECN) packets based on statistical probabilities. If the buffer is almost empty, all incoming packets are accepted. As the queue grows, the probability for dropping an incoming packet grows too. When the buffer is full, the probability has reached 1 and all incoming packets are dropped. RED is considered more fair than tail drop [2].

The more a host transmits, the more likely it is that its packets are dropped. Early detection helps avoid global synchronization. RED makes Quality of Service differentiation impossible. Weighted RED [5] and RED In/Out provide early detection with some QoS considerations. The Adaptive RED algorithm infers whether to make RED more or less aggressive based on the observation of the average queue length. If the average queue length oscillates around min threshold then Early Detection is too aggressive. On the other hand if the average queue length oscillates around max threshold then Early Detection is being too conservative. The algorithm changes the probability according to how aggressive it senses it has been discarding traffic.

RED has some disadvantages too, Unfairness is caused as nodes drop packets unaware of other nodes and Queue at any single node cannot reflect the network congestion state to overcome these problems Neighborhood Random Early Detection is proposed. It is based on the principle that Queue size of a neighborhood reflects the degree of local network congestion.

III. SYSTEM DESIGN - WIRELESS CONGESTION CONTROL PROTOCOL (WCCP)

TCP's congestion control suffers from a coarse granularity when applied to the multihop ad hoc environment. To overcome this problem a rate based wireless congestion control protocol is defined. There are two components in WCCP. One is at the transport layer. It replaces the window adjusting algorithm of TCP with a rate control algorithm to regulate the sending rate. The other is between the networking layer and the MAC layer. It monitors and possibly modifies the feedback field in TCP data packets when it passes the outgoing packets from the networking layer to the MAC layer and the incoming packets in the reverse direction.

A. System Architecture

Based on the estimate of the available bandwidth, the inter-node and intra-node resource allocation schemes are proposed to determine the available channel resource for each node and for each flow passing through that node and accordingly modify the MAC layer feedback. Then an end to- end rate control scheme is proposed to carry the feedback from the bottleneck node to the source node which accordingly adjust the sending rate to make full and fair utilization of the channel resource at the bottleneck node without causing severe medium contention and packet collision.

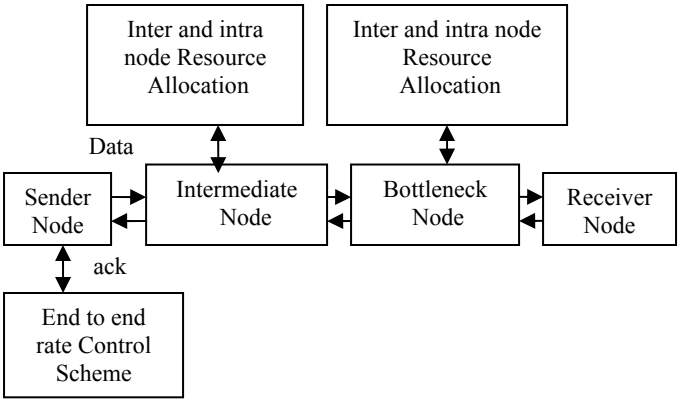


Figure 1. WCCP Mechanism

B. Measurement of Channel Busyness Ratio

The channel busyness ratio rb is an easily measured metric at the location of each node under the current architecture of the IEEE 802.11 standard. Notice that the IEEE 802.11 is a CSMA-based MAC protocol, working on the physical and virtual carrier sensing mechanisms. There is already a function to determine whether the channel is busy or not, i.e., the channel is determined busy when the measuring node is transmitting, receiving, or its network allocation vector indicates the channel is busy, and is idle otherwise. The channel busyness ratio can be simply calculated as the ratio of the total lengths of busy periods to the total time during a time interval, which is the average control interval.

C. Bandwidth Calculation

In the rate-based congestion control algorithm, to calculate the ideal sending rate, the source is in direct need of a timely and easily measured metric which should satisfy two requirements. First, since MAC contention is tightly coupled with congestion, a candidate of congestion signal should reflect the condition of MAC contention and collision. Second, in order to fully utilize the shared channel without causing severe congestion and packet collision, the candidate should indicate the available bandwidth. Channel busyness ratio rb, meets these two requirements. the channel utilization cu indicates the ratio of the channel occupancy time for successful transmissions to the total time, the normalized throughput s indicates the achievable data rate for payload divided by the channel data rate and is proportional to cu and the collision probability p indicates the probability that each transmission encounters a collision. the threshold thb should be chosen such that

$$rb \approx cu(rb \leq thb)$$

Since the performance of rb is not sensitive to n, we can fix n and observe the effect of the payload size. After choosing thb, according to Equation (3), we can estimate the available bandwidth of each node, denoted by BWa, as follows

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

$$BWa = BW\ (thb - rb)\ data/Ts,\ (rb<thb) \quad (1)$$

Where BW is the transmission rate in bits/s for the DATA packets, data is the average payload size in unit of channel occupancy time, and Ts is the average time of a successful transmission at the MAC layer.

D. Inter Node Resource Allocation

According to equation above equation, each node could calculate the total available bandwidth for its neighborhood based on the measured channel busyness ratio in a period called average control interval, denoted by Tc. To determine the available bandwidth for each node, WCCP accommodates the channel resource ΔS for each node proportionally to its current traffic load S in Tc.

$$\Delta S = (thb - rb)/rb * s \quad (2)$$

Because both the incoming traffic and outgoing traffic of each node consume the shared channel resource, S should include the total traffic (in bytes), i.e., the sum of the total incoming and outgoing traffic.There is two cases when we compare the observed rb with thb, i.e., rb <thb and rb ≥ thb. When rb < thb , ΔS is positive, meaning we should increase the traffic Since the available bandwidth is proportional to thb − rb according to equation (1), we may increase S by such an amount that after the increase ΔS, S is proportional to thb, which is the optimal channel utilization. Actually, equation (2) achieves our desired increase as it can be easily seen that

$$(\Delta S + S)/thb = s/rb \quad (3)$$

Therefore, rb will approach thb after one average control interval Tc when all the nodes in the neighborhood increase the total traffic rate according to equation (2). When rb ≥ thb, ΔS is negative, meaning we decrease the traffic. In this case, however, the linear relationship between the available bandwidth and rb no longer exists, and the collision probability increases dramatically as the total traffic rate increases. In addition, when the node enters saturation, both collision probability and rb amount to their maximum values and do not change as the traffic increases, although the total throughput decreases. It thus appears that ideally, WCCP needs to aggressively decrease the total traffic rate. However, since it is difficult to derive a simple relationship between the traffic rate and rb when rb ≥ thb, WCCP uses the same linear function as for the case rb < thb.. Indeed, this brings two advantages. First, as the increase and decrease use the same law, it is simple to implement at each node. Second, opting out of aggressive decrease helps achieve smaller oscillation in channel utilization.

E. Intra Node Resource Allocation

After calculating ΔS, the change in the total traffic or the aggregate feedback at each node, WCCP needs to apportion it to individual flows traversing that node in order to achieve both efficiency and fairness. WCCP relies on an Additive-Increase Multiplicative- Decrease policy to converge to efficiency and fairness: If ΔS > 0, all flows increase the same amount of throughput. And if ΔS < 0, each flow decreases the throughput proportionally to its current throughput.

Before determining the feedback when ΔS >= 0, WCCP needs to estimate the number of flows passing through the considered node. Again, since the channel is shared by both incoming and outgoing traffic,

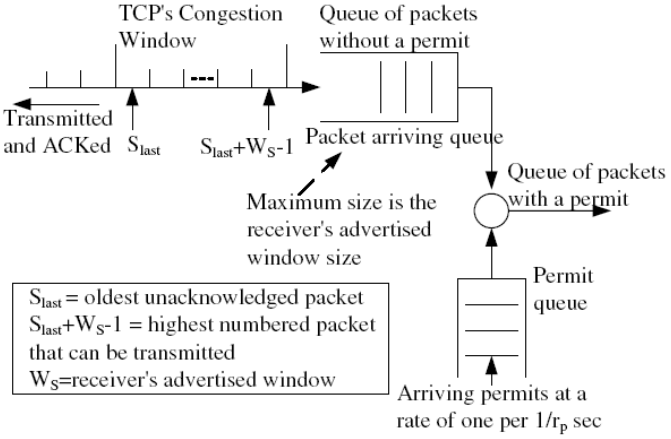


Figure 2. Rate control mechanism

The number of flows J used by WCCP should be different from the real number of flows. For those flows that either originate or terminate at the node, the node counts each as one flow, whereas for those flows only passing through the node, the node counts each as two flows, i.e., one in and one out. This is because that flows passing through the node occupy twice channel resource of that for flows originating or terminating at the node Let rpk denote the packet sending rate (pkt/s) of the flow which the kth observed packet during the period Tc at node i belongs to. Since rpkTc is equal to the number of packets which are observed from the corresponding flow at node I during Tc, we can convert the summation over each flow to the summation over each packet by the fact that, for all the packets of each flow, the summation of 1/rpkTc is equal to 1.Thus J can be calculated at node i as

$$J = 1/rph*Tc \quad \text{for k->1 to K} \quad (4)$$

Where K is the total number of packets seen by node i in Tc. Thus each node only needs to do the summation for each received and transmitted packet. Therefore, if ΔS >= 0, the increasing amount of traffic rate for each flow Cp, and per packet feedback pfk will be

$$Cp = \Delta S/Tcj \quad (5)$$

$$Pfk = Cn*Rpk/Rpk*Tc \quad (6)$$

where Cn is a constant and WCCP aims to make full use of the channel resource while not introducing severe medium contention, i.e., rb should be as close to thb as possible but never exceed thb too much. Therefore, when the aggregate feedback of previously passing packets is equal to ΔS, the node sets local feedback value as zero until the next control interval starts. With this mechanism in place, the channel busyness ratio rb should be around thb at the bottleneck nodes and be smaller at other nodes.

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

F. End-To-End Rate-Based Congestion Control Scheme

A leaky bucket (permit queue) is attached to the transport layer to control the sending rate of a WCCP sender. The permit arrival rate rp of the leaky bucket is dynamically adjusted according to the explicit feedback fb carried in the returned ACK whenever a new ACK arrives (henceforth, ACKs refer to the transport layer acknowledgments). Namely,

$$Rp=rp+fb \quad (7)$$

To enable this feedback mechanism, each WCCP packet carries a congestion header including three fields, i.e., rp, Tc, and fb, which is used to communicate a flow's state to the intermediate nodes and the feedback from the intermediate nodes to the source. The field rp is the sender's current permit arrival rate, and the field Tc is the sender's currently used control interval. They are filled in by the sender and never modified in transit. The last field, fb, is initiated by the sender and all the intermediate nodes along the path may modify it to directly control the packet sending rate of the source.

IV. SYSTEM IMPLEMENTATION

A. Tool Description

The network simulator ns-2 is an object–oriented, discrete event-driven network simulator developed at the OC Berkley and ISC ISI as part of the VINT project [VIN03]. It is a very useful tool for conducting network simulations involving local and wide area networks. In the recent years its functionality has grown to include wireless and ad hoc networks as well.

The ns-2 network simulator has gained an enormous popularity among participants of the research community, mainly because of its simplicity and modularity. The network simulation allows simulation scripts, also called simulation scenarios, to be easy written in a script-like programming language TCL. Emulation refers to the ability to introduce the simulator into a live network. Special objects within the simulator are capable of introducing live traffic into the simulator and injecting traffic from the simulator into the live network. The interface between the simulator and live network is provided by a collection of objects including Tap Agents and Network Objects. Tap agents embed live network data into simulated packets and vice-versa. Network objects are installed in tap agents and provide an entry point for the sending and receipt of live data. Both objects are described in the following sections. When using the emulation mode, a special version of the system scheduler is used: the Real Time Scheduler. This scheduler uses the same underlying structure as the standard calendar-queue based scheduler, but ties the execution of events to real time

More complex functionality relies on C++ code that either comes with ns-2 or is supplied by the user. The utilization of the two programming languages increases the flexibility of ns-2. C++ is mainly used for event handling and per-packet processing tasks for which TCL would become too slow. TCL is most commonly used for simpler routing protocols, general ns-2 code and simulation scenario scripts. The usage of TCL for simulation scenario scripts allows the user to change parameters of a simulation without having to recompile any source code.

Simulations in ns-2 can be logged to trace files, which include detailed information about packets in the simulation and allows for post-run processing with some analysis tool. It is also possible to let ns-2 generate a special trace file that can be used by NAM (Network Animator), a visualization tool that is part of the simulator distribution.

B. Resource Requirements

- Hardware requirements

Processor Type	:	Pentium-IV,512MB RAM
Hard Disk	:	20 GB

- Software requirements

Operating System	:	Linux
Programming Language	:	C++
Tool	:	Network Simulator 2

V. RESULTS AND DISCUSSION

A. NAM Output

The nam output shows the 9-node chain topology. There are three cases considered to compare the performance of TCP and Wireless congestion Control Protocol (WCCP).In case 1 there is only one flow between node0 to node 8 and in case 2 there are three flows one flow is between node0 to node4 and the second flow is between node0 to node8 and third flow is between node4 to node8.

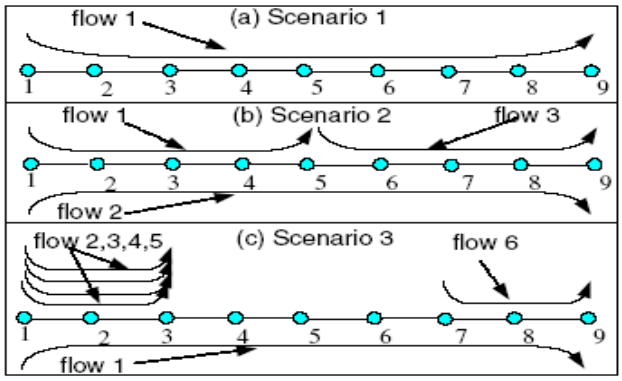


Figure 3. 9-Node Chain Topology with Different Traffic Distribution

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

Figure 3 shows various flows for the nine node topology. Scenario 1 has only one flow from node 1 to node9 and Scenario 2 has 3 flows and Scenario 3 has 6flows

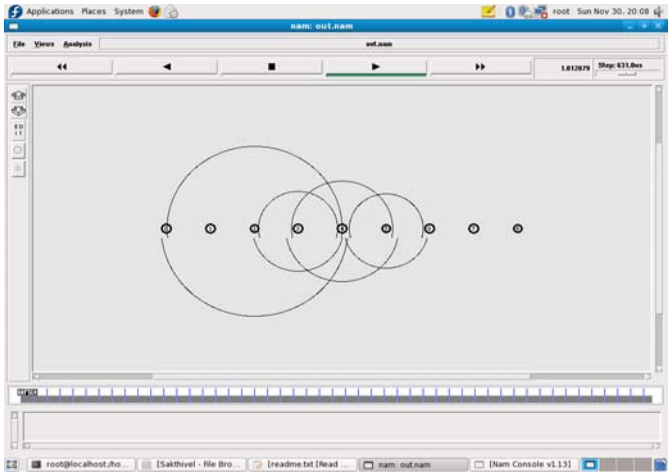


Figure 4. Nam Output

Figure 4 shows the Nam Output for the nine node topology as shown in the figure 3 it shows the Nam output for 3 flows

B. Graphs

1) Scenario1

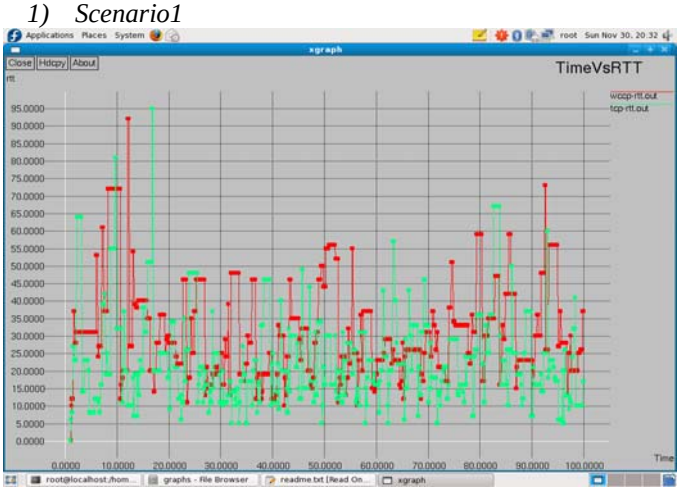


Figure 5. Round Trip Time of TCP vs WCCP

Figure 5. Shows the comparison of round trip time of TCP vs WCCP for scenario1

2) Scenario 2

The graph shows the comparison of throughput for TCP and WCCP. Throughput of WCCP is higher than TCP.

Figure 6 shows the throughput of TCP for various flows. Here there are three flows. Throughput of the second flow which is the main flow is much smaller than the other flows.

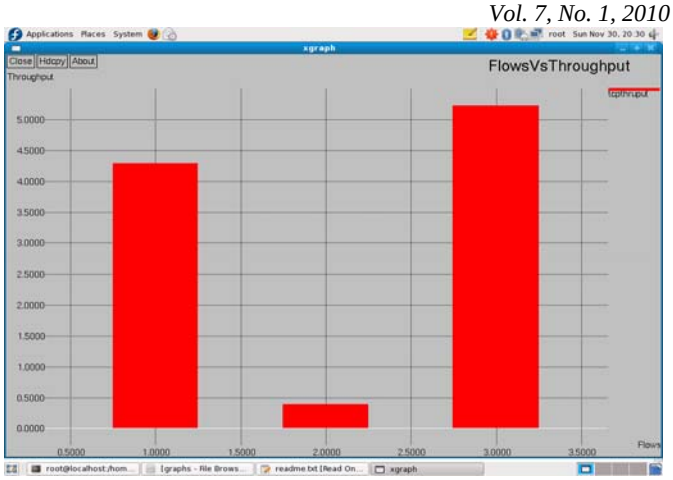


Figure 6. TCP Throughput

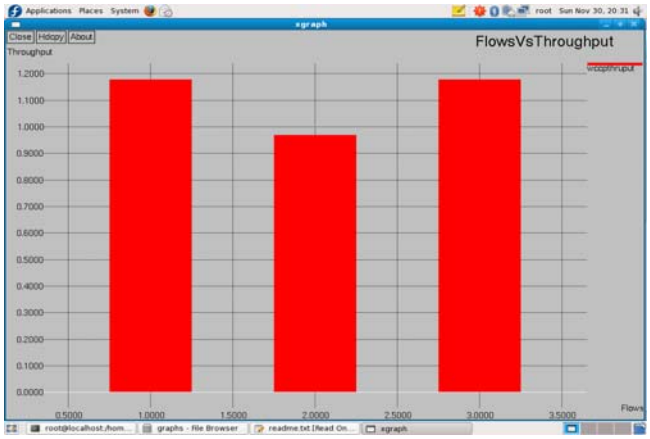


Figure 7. WCCP Throughput

Figure 7 shows the throughput of WCCP for various flows. Here there are three flows. Throughput for the second flow is very small for TCP but it is overcome by Wireless Congestion Control Protocol. Here it overcomes the problem by allocating the resources properly at each node and by providing the required feedback to the source node.

3) Scenario 3

The graph shows the comparison of throughput for TCP and WCCP. Throughput of WCCP is higher than TCP.

Figure 8 shows the throughput of TCP for various flows. Here there are six flows. Four flows from node1 to node3 and one from node7 to node9 and the main flow is from node1 to node9 and due to many flows the throughput for main flow is zero

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

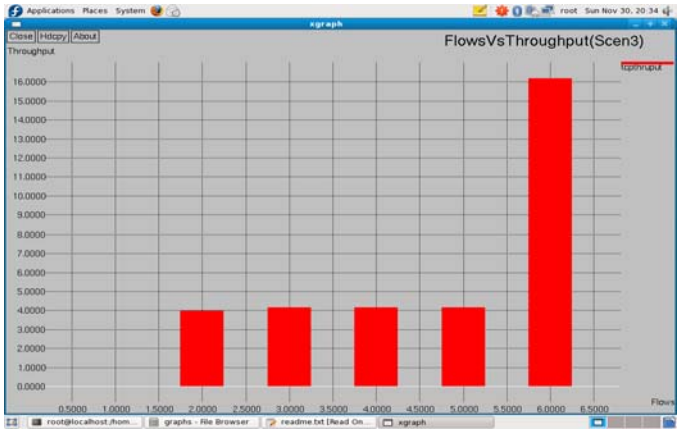


Figure 8. TCP Throughput

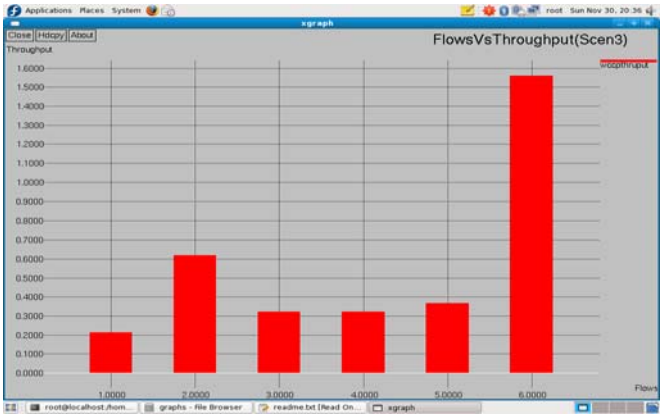


Figure 9. WCCP Throughput

Figure 9 shows the throughput of TCP for various flows. Here there are six flows. Four flows from node1 to node3 and one from node7 to node9 and the main flow is from node1 to node9 and due to many flows the throughput for main flow is zero in TCP but it is overcome by WCCP.

VI. CONCLUSIONS AND FUTURE ENHANCEMENT

Congestion control is critical to reliable transport service in wireless multihop ad hoc networks. Unfortunately, traditional TCP suffers severe performance degradation and unfairness. Realizing that the main reason is the poor interaction between traditional TCP and the MAC layer, we propose a systematic solution named Wireless Congestion Control Protocol (WCCP) to address this problem in both layers. WCCP uses channel busyness ratio to allocate the shared resource and accordingly adjusts the sender's rate so that the channel capacity can be fully utilized and fairness is improved. We evaluate WCCP in comparison with TCP in various scenarios. The results show that our scheme outperforms traditional TCP in terms of channel utilization, end-to-end delay, and fairness, and solves the starvation problem of TCP flows.

REFERENCES

[1] [1] X. Chen, H. Zhai, J. Wang, and Y. Fang, "TCP performance over mobile ad hoc networks," Canadian J. Electric. Comput. Engin., vol. 29, no. 1/2, pp. 129-134, Jan./April 2004.

[2] H. Zhai, X. Chen, and Y. Fang, "Rate-based transport control for mobile ad hoc networks," in Proc. IEEE WCNC., vol. 8, no. 11, pp. 12-19, march 2005.

[3] S. Xu and T. Safadawi, "Does the IEEE 802.11 MAC protocol work well in multihop wireless ad hoc networks? IEEE Commun. Mag., vol. 39, no. 6, pp. 130–137, June 2001.

[4] Z. Fu, X. Meng, and S. Lu, "How bad TCP can perform in mobile ad-hoc networks," in Proc. IEEE Symposium on Computers and Communications, vol. 6, no. 2, pp. 219-228 2002.

[5] H. Zhai, J. Wang, X. Chen, and Y. Fang, "Medium access control in mobile ad hoc networks: challenges and solutions," Wireless Commun. Mobile Comput., vol. 6, no. 2, pp. 151–170, March 2006.

[6] Z. Fu, P.Zerfos, H. Luo, S. Lu, L. Zhang, and M. Gerla, "The impact of multihop wireless channel on TCP throughput and loss," in Proc. IEEE INFOCOM.,vol 3, no 6 pp. 1744-1753, April 2003.

[7] K. Chen, Y. Xue, and K. Nahrstedt, "On setting TCP's congestion window limit in mobile ad hoc networks," in Proc. IEEE ICC, vol 6, no 11, pp.153-167, May 2005.

[8] T. D. Dyer and R. V. Boppana, "A comparison of TCP performance over three routing protocols for mobile ad hoc networks," in Proc. ACM Mobihoc ,vol 6, no 2, pp.662-678, June 2004.

[9] J. P. Monks, P. Sinha, and V. Bharghavan, "Limitations of TCP-ELFN for ad hoc networks," in Proc. MOMUC, vol 8, no 11, pp.47-67, March 2002.

[10] X. Yang and N. Vaidya, "On physical carrier sensing in wireless ad hoc networks," in Proc. IEEE INFOCOM , vol 11, no 5, pp.603-615, May 2005.

AUTHORS PROFILE

Mahendra Kumar.S received the Bachelors degree in Electronics and Communication Engineering from from Anna University, Chennai in 2006. and the Master degree in Communication System from Kumaraguru College of Technology, Coimbatore in 2009. He is currently working in Velalar College of Engineering and Technology as Lecturer of Electronics and Communication Department since 2009. His fields of interests include Ad Hoc & Sensor Network, He has published two papers in National Conferences and one International Conference in Ad Hoc Networks. He is a life member of ISTE.

Senthil Prakash.K received the Bachelors degree in Electronics and Communication Engineering from Bharathidasan University, Tiruchirappalli in 2003. and the Master degree in Communication System from Kumaraguru College of Technology, Coimbatore in 2009. He is currently working in Velalar College of Engineering and Technology as Lecturer of Electronics and Communication Department since 2009. His area of interest includes Mobile communication and Ad Hoc Network, He has published two papers in National Conferences in Ad Hoc Networks. He is a life member of ISTE.

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

Saturation Throughput Analysis of IEEE 802.11b Wireless Local Area Networks under High Interference considering Capture Effects

Ponnusamy Kumar
Department of Electronics and Communication Engineering
K.S.Rangasamy College of Technology
Tiruchengode, Namakkal, Tamilnadu, India

A.Krishnan
Department of Electronics and Communication Engineering
K.S.Rangasamy College of Technology
Tiruchengode, Namakkal, Tamilnadu, India

***Abstract*— Distributed contention based Medium Access Control (MAC) protocols are the fundamental components for IEEE 802.11 based Wireless Local Area Networks (WLANs). Contention windows (CW) change dynamically to adapt to the current contention level: Upon each packet collision, a station doubles its CW to reduce further collision of packets. IEEE 802.11 Distributed Coordination Function (DCF) suffers from a common problem in erroneous channel. They cannot distinguish noise lost packets from collision lost packets. In both situations a station does not receive its ACK and doubles the CW to reduce further packet collisions. This increases backoff overhead unnecessarily in addition to the noise lost packets, reduces the throughput significantly. Furthermore, the aggregate throughput of a practical WLAN strongly depends on the channel conditions. In real radio environment, the received signal power at the access point from a station is subjected to deterministic path loss, shadowing and fast multipath fading. In this paper, we propose a new saturation throughput analysis for IEEE 802.11 DCF considering erroneous channel and capture effects. To alleviate the low performance of IEEE 802.11 DCF, we introduce a mechanism that greatly outperforms under noisy environment with low network traffic and compare their performances to the existing standards. We extend the multidimensional Markov chain model initially proposed by Bianchi[3] to characterize the behavior of DCF in order to account both real channel conditions and capture effects, especially in a high interference radio environment.**

Keywords-throughput; IEEE802.11; MAC; DCF; capture

I. INTRODUCTION

The use of IEEE 802.11 wireless local area networks (WLANs) has been spreading quickly. One of the channel access mechanisms in IEEE 802.11 is Distributed Coordination Function (DCF). DCF is based on Carrier Sense Multiple Access with Collision Avoidance(CSMA/CA) algorithm. In this mechanism, a station waits for a quiet period in wireless media, and then begins to transmit data while detecting collisions. The time lapse between successive carrier senses, when channel is occupied, is given by a back-off counter which has an initial random value within a predetermined range. The standard also defines an optional access method, PCF, which is for time bounded traffic. Since the DCF mechanism has been widely adopted in wireless networks, we focus our analysis only on this mechanism.

Many research efforts have been done to study the IEEE 802.11 DCF performance, by both of analysis and simulation. Most of them assume the ideal channel condition, which means that the packet corruptions are only due to collisions[3-5]. A detailed analysis for the multi-access behavior in the 802.11 DCF is presented in [6], where the packet sending probability p, which depends on different contention window size, is computed by approximating the 802.11 DCF under saturated traffic as a p-persistent CSMA protocol. This approximation is very useful for the analysis of the DCF and several papers [7-10] have adopted this approach and analyzed saturated DCF performance. Bianchi [3] suggests a Markov model to represent the exponential backoff process, and Wu et al. [8] use the same model and take the packet retry limit into account.

In [11] , based on the IEEE 802.11 DCF, a novel scheme named DCF is proposed to improve the performance of Wireless Local Area Network (WLAN) in fading channel. Impact of bursty error rates on the performance of wireless local area network is studied in [12]. The throughput and delay were analyzed in ideal and error-prone channels.

In [13], the authors proposed a fast collision resolution (FCR) algorithm. In this algorithm, when a station detects a busy period, it exponentially increases its contention window and generates a new backoff counter. In case that a station detects a number of consecutive idle slots, it exponentially reduces the backoff counter. In [14], the authors proposed a new backoff algorithm to measure the saturation throughput under several conditions and several set of parameters which are adjusted dynamically according to the network conditions.

In [15], the authors presented an extension of Bianchi‘s model to a non saturated environment. They modified the multi-dimensional Markovian state transition model by including state, characterizing the system when there are no packets to be transmitted in the buffer of a station. These states are called post backoff states and denote a kind of virtual backoff counter initiated prior to packet arrival. In [16], the authors propose a new scheme for IEEE 802.11 DCF to alleviate the low performance of the high date rate stations for

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

asynchronous networks. They introduce an adaptive mechanism to adjust the packet size according to the data rate, in which the stations occupy the channel equal amount of time.

In real radio environment, the signal power at access point from a given station will be subjected to deterministic path loss, shadowing and fast multi-path fading. Due to this, when more than one station simultaneously transmit to access point, the channel is successfully captured by a station whose signal power level is stronger than the other stations and thus increases the actual throughput. This phenomenon is called capture effect. In [17], the authors presented a Markov model to analyze the throughput of IEEE802.11 considering transmission errors and capture effects over Rayleigh fading channels in saturated network conditions. Their model is very accurate when the contention level of a network is high. In [18], we have presented a novel scheme for DCF under non saturated traffic condition. In [19], we extend [3] and presented the non saturation throughput analysis for heterogeneous traffic. Hadzi-velkov and Spasenovski [20] have investigated the impact of capture effect on IEEE 802.11 basic service set under the influence of Rayleigh fading and near/far effect.

Liaw et al. [22] introduced an idle state, not present in Bianchi's model [3], accounting for the case in which the station buffer is empty after a successful completion of a packet transmission. The probability that there is at least a packet in the buffer after a successful transmission is assumed to be constant and independent of the access delay of the transmitted packet. In [23], we presented the performance study for multihop network in string topology.

In this paper, we present an analytical model to study the saturation behavior of the IEEE802.11 DCF considering erroneous channel and capture effects. We differentiate channel induced errors from packet collision in order to optimize the performance of CSMA/CA under the saturated network condition.

The rest of the paper is organized as follows: Section II describes our model for basic access mechanism under saturated traffic condition. Performance of the proposal scheme is analyzed in section III. Finally, section VI concludes this paper.

II. Performance Analysis For 802.11 DCF In Saturated Trafic Condition

In this section, we present a discrete time bi-dimensional Markov model for evaluating the saturation throughput of the DCF under non ideal channel conditions considering capture effects. Saturated traffic condition means that all users always have a packet available for transmission. Throughput under saturated traffic situation is the upper limit of the throughput achieved by the system, and it represents the maximum load the system can carry in the stable condition.

Let process s(t) be the stochastic process representing the backoff stage of a given station at the given time t. A second process b(t) is defined, representing backoff time counter of the station. Backoff time counter is decremented at the start of every idle backoff slot. The backoff counter is an integer value uniformly chosen from $[0,W_i-1]$ where W_i denotes the contention window at the i^{th} backoff stage. The backoff stage 'i' is incremented by one for each failed transmission attempt up to the maximum value m, while the contention window is doubled for each backoff stage up to the maximum value $W_{max} = 2^m W_{min}$.

Letting $W_{min} = W_0$, we can summarize the W as,

$$W_i = \begin{Bmatrix} 2^i W_0, 0 \le i \le m \\ 2^m W_0, \ i > m \end{Bmatrix} \quad (1)$$

The main aim in this section is to modify the MAC protocol in order to enhance the performance of MAC protocol in the event of channel induced errors. The assumption that all frame losses are due to collisions between WLAN devices is generally not true in a noisy wireless environment. However, unsuccessful reception of the data frame can also be caused by channel noise or other interference.

In case of unsuccessful transmission the basic BEB mechanism will double the contention window size by considering channel errors as a packet collision. This process will unnecessarily increase the backoff overhead and intern increases channel idle slots. In order to alleviate this problem we propose a new mechanism that takes advantage of a new capability to differentiate the losses, and thereby sharpen the accuracy of the contention resolution process. When the frame is corrupted by the channel induced noise, we maintain the same contention size instead of doubling it.

A. Loss differentiation method for basic access mechanism

Basic access mechanism is the default access method in DCF and employs a two-way handshaking procedure. The loss differentiation for basic access is not straightforward because it provides only ACK feedback from receivers. The loss differentiation method for WLAN has been proposed in [24]. The following describes a loss differentiation method for basic access which requires minimum modifications to the legacy standard to provide additional feedback. The data frame can be functionally partitioned into two parts: header and body. The MAC header contains information such as frame type, source address and destination address, and comes before the MAC body, which contains the data payload.

In a WLAN with multiple stations sharing a common channel, a collision occurs when two or more stations starts transmission in the same time slot, which will likely corrupt the whole frame (header plus body) at the receiver end. On the other hand, a frame transmission that is not affected by collision with transmission from another WLAN station may still be corrupted by noise and interference. However, under the condition that the signal-to-noise-plus-interference ratio (SINR) is reasonable to maintain a connection between the sending and receiving stations, the receiver is likely able to acquire the whole data frame and decode it, as the physical header is transmitted at the base data rate for robustness (e.g., in 802.11b, the 192-bit physical header is always transmitted at 1Mbps). In this case, the noise or interference may result in a few bit-errors that cause a Frame Check Sequence (FCS) error in the decoded data frame, which is then discarded by the receiver station. As the MAC header (18-30 bytes) is typically much shorter than the MAC body (e.g., a typical

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

Internet Protocol datagram is several hundreds to a couple of thousands bytes long), when FCS fails, it is much more likely caused by bit errors in the body than the header.

If the MAC header is correctly received but the body is corrupted, the receiver can observe the MAC header content to learn the identity of the sender and to verify that it is the intended receiver. To verify the correctness of the MAC header, a short Header Error Check (HEC) field can be added at the end of the header as shown in Fig.1 in order to provide error checking over the header, while the FCS at the end of the frame provides error checking over the entire MAC frame. Note that the use of HEC in the header is not a new concept as it has been adopted in many other communication systems, such as asynchronous transfer mode and Bluetooth, all of which includes a 1-byte HEC or header check sequence (HCS) field in their header. With the HEC, when a data frame is received and FCS fails, the HEC can be verified to see if the header is free of error, and if so, proper feedback can be returned to the sender identified by the MAC header.

As discussed above, FCS failure but correct HEC in a frame reception is a good indication that the frame has been corrupted by transmission errors rather than a collision. Because in the basic access mechanism, only ACK frames are available to provide positive feedback, a new control frame NAK needs to be introduced to inform the sender that the data frame transmission has failed and the failure is due to transmission errors; i.e., the data frame has suffered a transmission loss. On the other hand, if the sender receives neither a NAK nor an ACK after sending a data frame, it is a good indication that the frame transmission has suffered a collision loss. The NAK frame can be implemented with exactly the same structure as the ACK frame except for a one-bit difference in the frame type field in the header, and is sent at the same data rate as an ACK frame. The transmission of a NAK does not consume more bandwidth or collide with other frames because it is transmitted SIFS after the data frame transmission and occupies the time that would have been used by the transmission of an ACK.

The HEC field is a necessary modification to the standard because without it, when the FCS fails, the receiver would not be able to determine if the header is in error and would not be able to trust the sender address in the header for returning the NAK. The HEC field (which can be 1 or 2 bytes) costs an extra overhead. But it can be calculated that the overhead due to the extra field to the total transmission time is much less than 1%. Therefore the overhead is negligible. Comparing the two loss differentiation methods, RTS/CTS access is useful when the data frame size or the number of stations is very large or there are hidden terminals. However, as it consumes extra time for RTS/CTS exchange, RTS/CTS access is less efficient than basic access in other cases.

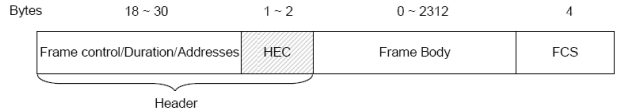


Figure 1. Frame format for Basic access mechanism

The loss differentiation method for RTS/CTS access does not involve any modification to the standard. The loss differentiation method for basic access, however, needs two minor modifications to the current standard: the HEC field and the NAK frame, which are both easy to implement.

B. Analitical modeling of new backoff algorithm

Based on the above consideration, let us discuss the Markovian model shown in Fig.2, assuming saturation network conditions. We assume that each station has m+1 stages of backoff process. The value of the backoff counter is uniformly chosen in the range $(0,W_i\text{-}1)$, where $W_i = 2^i W_{min}$ and depend on the station's backoff stage *i*. A station in (i,0) state will transit into (i+1,k) state in the event of collision without capture effect. On the other hand, the model transits from (i,0) to (0,k) state if frame is successfully captured. From state (i,0), the station re-enters the same backoff stage (i,k) in case of unsuccessful transmission due to transmission errors.

The main approximation in our model is that, at each transmission attempt, each packet collides with constant and independent probability P_{col} regardless of previously suffered attempts and transmission errors occur with probability P_e due to the erroneous channel. We also assume that the channel is captured by a station with the probability P_{cap} in the event of collision. Based on the above assumptions we can derive the transition probabilities:

$$
\begin{aligned}
&P\{i,k/i,k+1\} = 1, \;\; k \in [0, W_i - 2], \;\; i \in [0,m] \\
&P\{0,k/i,0\} = ((1-P_{col})(1-P_e) + P_{col}P_{cap})/W_0, \;\; k \in [0, W_i - 1], \;\; i \in [0,m] \\
&P\{i,k/i,0\} = (1-P_{col})P_e/W_0, \;\; k \in [0, W_i - 1], \;\; i \in [0,m] \\
&P\{i,k/i-1,0\} = (1-P_{cap})P_{col}/W_i, \;\; k \in [0, W_i - 1], \;\; i \in [1,m] \\
&P\{m,k/m,0\} = ((1-P_{cap})P_{col} + (1-P_{col})P_e)/W_m, \;\; k \in [0, W_m - 1]
\end{aligned} \qquad (2)
$$

The first equation represents that, at the beginning of each time slot, the backoff time is decremented. The second equation states that, the initialization of backoff window after successful transmission for a new packet. The third equation accounts that, the maintenance of backoff window in the same stage, if channel error is detected. The fourth and fifth equations represent that, the rescheduling of backoff stage after unsuccessful transmission.

Let the stationary distribution of the chain be $b_{i,k} = \lim_{t->\infty} P\{s(t) = i, b(t) = k\}, i \in (0,m), k \in (0, W_{i-1})$. To obtain the closed form solution we first consider the following relations:

$$
\begin{aligned}
b_{i,0} &= b_{i-1,0}\{P_{col}(1-P_{cap})\} + b_{i,0}\{P_e(1-P_{col})\} \\
&= \left(\frac{P_{col}(1-P_{cap})}{1-(1-P_{col})P_e}\right) b_{i-1,0} \\
&= \left(\frac{P_{col}(1-P_{cap})}{1-(1-P_{col})P_e}\right)^i b_{0,0} \quad 0 < i < m
\end{aligned} \qquad (3)
$$

and,

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

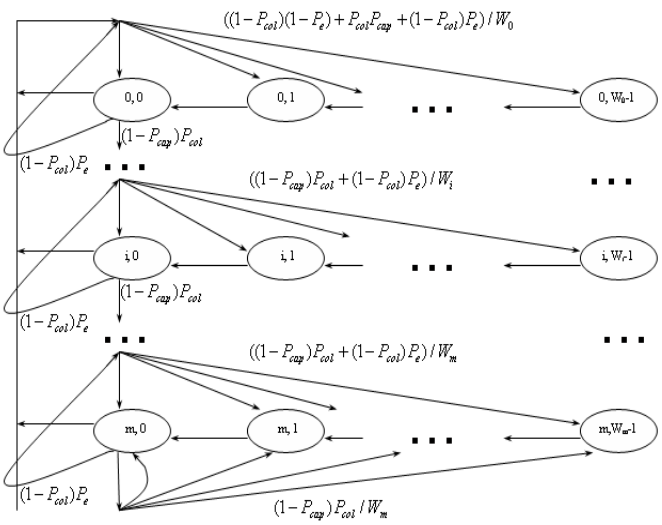


Figure 2. Markov chain model for the backoff procedure of a station

$$b_{m,0} = b_{m-1,0}(1-P_{cap})P_{col} + b_{m,0}P_{col}(1-P_{cap}) + b_{m,0}(1-P_{col})P_e \quad (4)$$

$$b_{m,0}\left\{1-\left(\frac{P_{col}(1-P_{cap})}{1-(1-P_{col})P_e}\right)\right\} = b_{m-1,0}\left(\frac{P_{col}(1-P_{cap})}{1-(1-P_{col})P_e}\right) \quad (5)$$

from which we obtain the following relation,

$$b_{m,0} = \frac{\left(\frac{P_{col}(1-P_{cap})}{1-(1-P_{col})P_e}\right)^m b_{0,0}}{1-\left(\frac{P_{col}(1-P_{cap})}{1-(1-P_{col})P_e}\right)} \quad (6)$$

A closed-form solution to the Markov chain owing to the chain regularities, for each k ϵ (1,W_i-1), shown as:

$$b_{i,k} = \frac{W_{i-k}}{W_i}\begin{cases} ((1-P_{col})(1-P_e)+P_{col}P_{cap})\sum_{i=0}^{m} b_{i,0} + b_{i,0}(1-P_{col})P_e, & i=0 \\ P_{col}(1-P_{cap})b_{i-1,0} + (1-P_{col})P_e b_{i,0}, & 1 \le i \le m \\ (P_{col}(1-P_{cap}))(b_{m-1.0}+b_{m,0}) + (1-P_{col})P_e b_{m,0}, & i=m \end{cases} \quad (7)$$

By means of relations (3), (5) and remembering

$$\sum_{i=0}^{m} b_{i,0}\left(1-\frac{P_{col}(1-P_{cap})}{1-(1-P_{col})P_e}\right) = b_{0,0}$$

we rewrite the relation (7) as:

$$b_{i,k} = \frac{W_i - k}{W_i} b_{i,0} \quad i \in (0,m),\ k \in (0, W_i - 1) \quad (8)$$

Thus, by relations (3), (5) and (8), all the values of $b_{i,k}$ are expressed as a function of $b_{0,0}$. Considering normalization conditions, and making use of the above equations we obtain the following relation:

$$\sum_{i=0}^{m}\sum_{k=0}^{Wi-1} b_{i,k} = 1$$

$$= \frac{b_{0,0}}{2}\left[W_0\left(\sum_{i=0}^{m-1}(2P_t)^i + \frac{(2P_t)^m}{1-P_t}\right) + \frac{1}{1-P_t}\right]$$

from which, we obtain,

$$b_{0,0} = \frac{2(1-P_t)(1-2P_t)}{W_0(1-P_t)(1-(2P_t)^m) + W_0(2P_t)^m(1-2P_t) + (1-2P_t)} \quad (9)$$

where we assume,

$$P_t = \frac{P_{col}(1-P_{cap})}{1-(1-P_{col})P_e}$$

Assuming error free channel with no capture effects, i.e., $P_e = P_{cap} = 0$, then (9) can be rewritten as,

$$b_{0,0} = \frac{2(1-2P_{col})(1-P_{col})}{(W+1)(1-2P_{col}) + W(1-(2P_{col})^m)P_{col}} \quad (10)$$

which is similar to $b_{0,0}$ found in Bianchi's model[3] under saturated load conditions.

Now we can express the probability τ that a station transmits in a randomly chosen slot time when the backoff time is zero as,

$$\tau = \sum_{i=0}^{m} b_{i,0} = \frac{b_{0,0}}{1-P_t} \quad (11)$$

By substituting (9) in (11), we obtain the following relation.

$$\tau = \frac{2(1-2P_t)}{W_0(1-P_t)(1-(2P_t)^m) + W_0(2P_t)^m(1-2P_t) + (1-2P_t)} \quad (12)$$

Note that, when m=0, that is no exponential backoff is considered, and assuming $P_{cap}=P_e=0$, the probability τ results to be independent of collision probability under saturated traffic condition

$$\tau = \frac{2}{W_0 + 1} \quad (13)$$

which is the result found in[3] for constant backoff window.

However, in general, the probability τ depends on the conditional collision probability P_{col}, capture probability P_{cap} and probability of packet loss P_e. In our model we assume basic access method to compute the conditional collision probability P_{col}. To determine the value P_{col} it is sufficient to note that the probability that a transmitted packet encounters a collision if in a given time slot, at least one of the remaining (n-1) stations transmits another packet simultaneously. The conditional collision probability also depends on the capture probability because capture effect is the sub event of collision, i.e. without collision there is no capture effect. Therefore the probability P_{col} can be expressed as,

$$P_{col} = 1-(1-\tau)^{n-1} - P_{cap} \quad (14)$$

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

Our proposed model considers deterministic power loss and multipath fast fading of transmitted signals into account. We also assume that there is no direct path between transmitter and receiver within Basic Service Set(BSS), which means the envelop of received signal is Rayleigh faded. To compute the capture probability, we use the model proposed by Hadzi-velkov and Spasenovski[10]. In Rayleigh fading channel, the transmitted instantaneous power is exponentially distributed according to

$$f(p) = \frac{1}{p_0}\exp(-\frac{p}{p_0}), \qquad p > 0 \tag{15}$$

where p_0 represent the local mean power of the transmitted frame at the receiver and is determined by

$$p_0 = A.r_i^{-x}.p_t$$

where r_i is the mutual distance from transmitter to receiver, x is the path loss exponent, $A.r_i^{-x}$ is the deterministic path loss and p_t is the transmitted signal power. The path loss exponent for indoor channels in picocells is typically taken as 4. During simultaneous transmission of multiple stations, a receiver captures a frame if the power of detected frame p_d sufficiently exceeds the joint power of 'n' interfering contenders

$$p_{\text{int}} = \sum_{k=1}^{n} p_k$$

by a certain threshold factor for the duration of a certain time period. Thus capture probability is the probability of signal to interference ratio

$$\gamma = \frac{p_d}{p_{\text{int}}} \tag{16}$$

exceeding the product $z_0 g(S_f)$ where z_0 is known as the capture ratio and $g(S_f)$ is processing gain of the correlation receiver. The processing gain introduces a reduction of interference power by a factor $g(S_f)$, which is inversely proportional to spreading factor S_f. The conditional capture probability P_{cap} can be expressed over i interfering frames as,

$$\begin{aligned} P_{cap}(z_0 g(S_f) \mid i) &= prob(\gamma > z_0 g(S_f)/i) \\ &= [1 + z_0 g(S_f)]^{-i} \end{aligned} \tag{17}$$

For Direct Sequence Spread Spectrum(DSSS) using 11 chip spreading factor (s_f=11),

$$g(S_f) = \frac{2}{3S_f}$$

Now the frame capture probability can be expressed as,

$$P_{cap}(z_0, n) = \sum_{i=1}^{n-1} R_i P_{cap}(z_0 g(S_f) \mid i) \tag{18}$$

where R_i is the probability of 'i' interfering frames being generated in the generic time slot, according to

$$R_i = \binom{n}{i+1} \tau^{i+1} (1-\tau)^{n-i-1} \tag{19}$$

Next step is the computation of the saturation system throughput, defined as the fraction of time the channel is used successfully to transmit the bits. Let P_{tr} be the probability that there is at least one transmission in the considered time slot, with n stations contending for the channel, and each transmits with probability τ,

$$P_{tr} = 1 - (1-\tau)^n \tag{20}$$

The probability P_s that a transmission on the channel is successful is given by the probability that exactly one station transmit on the channel or probability that two or more stations transmit simultaneously where one station captures the channel due to capture effects,

$$P_s = \frac{n\tau(1-\tau)^{n-1} + P_{cap}}{1-(1-\tau)^n} \tag{21}$$

Now we can express throughput as,

$$S = \frac{E[\,payload\ transmitted\ in\ a\ timeslot\,]}{E[\,length\,of\ a\ timeslot\,]}$$

$$= \frac{P_{tr}P_s(1-P_e)E[PL]}{(1-P_{tr})\sigma + P_{tr}(1-P_s)T_c + P_{tr}P_sP_eT_e + P_{tr}P_s(1-P_e)T_s} \tag{22}$$

where, T_c is the average time that the channel sensed busy due to collision, T_s is the average time that the channel sensed busy due to successful transmission, T_e is the average time that the channel is occupied with error affected data frame and σ is the empty time slot. For the basic access method we can express the above terms as,

T_c= H + E[PL] + $ACK_{timeout}$

T_s= H + E[PL] + SIFS + ACK + DIFS + $2\tau_d$

T_e= H + E[PL] + NAK

Here, H – Physical header + MAC header

E[PL] – Average payload length and

τ_d – propagation delay

III. PERFORMANCE EVALUATIONS

In what follows, we shall present the results for the data rate of 11Mbps. In the results presented below we assume the following values for the contention window: W_{min}=32, m=5 and W_{max}=1024. The network parameters of 802.11b are given in Table I. We have also examined 802.11b with other possible parameter values. We use the method given in the IEEE standards [2] to calculate the bit error rates (BERs) and frame error rates (FERs) in a WLAN. This method has also been used in [25]- [27] to study WLAN performance. It is briefly described as follows.

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

TABLE I. NETWORK PARAMETERS

MAC header	24 bytes
PHY header	16 bytes
Payload size	1024 bytes
ACK	14 bytes
NAK	14 bytes
Basic rate	1Mbps
Data rate	11Mbps
τ_d	1 μs
Slot time	20 μs
SIFS	10 μs
DIFS	50 μs
ACK timeout	300 μs

A. Bit error rate (BER) model for 802.11b

First, the symbol error rate (SER) is calculated based on the signal-to-noise-plus-interference ratio (SINR) at the receiver. It is assumed that the interference and noise affect the desired signal in a manner equivalent to additive white Gaussian noise (AWGN). Given the number of bits per symbol, the SER is then converted into an effective BER. IEEE 802.11b uses DBPSK modulation for basic data rate at 1Mbps and complementary code keying (CCK) modulation to achieve its higher data rates (5.5 Mbps and 11 Mbps). The SER in CCK [2] has been determined as:

$$SER = \sum Q(\sqrt{2 \times SINR \times R_c \times D_c}) \tag{23}$$

where, R_c is the code rate, D_c is the codeword distance, and, Σ is over all codewords. For 11 Mbps data rate, the SER is given by

$$\begin{aligned} SER_{11Mbps} = {} & 24 \times Q(\sqrt{4 \times SINR}) + 16 \times Q(\sqrt{6 \times SINR}) \\ & + 174(\sqrt{8 \times SINR}) + 16 \times Q(\sqrt{10 \times SINR}) \\ & + 24 \times Q(\sqrt{12 \times SINR}) + Q(\sqrt{4 \times SINR}) \end{aligned} \tag{24}$$

As each symbol encodes 8 bits in 11 Mbps, the BER is

$$BER_{11Mbps} = \frac{2^{8-1}}{2^8 - 1} \times SER_{11Mbps} = \frac{128}{255} \times SER_{11Mbps} \tag{25}$$

The SER for 5.5 Mbps is calculated as,

$$SER_{5.5Mbps} = 14 \times Q(\sqrt{8 \times SINR}) + Q(\sqrt{16 \times SINR}) \tag{26}$$

And

$$BER_{5.5Mbps} = \frac{2^{4-1}}{2^4 - 1} \times SER_{11Mbps} = \frac{8}{15} \times SER_{11Mbps} \tag{27}$$

The SER in DBPSK modulation scheme has been determined as:

$$SER_{1Mbps} = Q(\sqrt{11 \times SINR}) \tag{28}$$

For 1Mbps mode, because each symbol encodes a single bit, the BER is the same as SER. In case of 2Mbps, the BER is calculated as,

$$SER_{2Mbps} = Q(\sqrt{5.5 \times SINR}) \tag{29}$$

When the BERs have been determined, the FERs of the data and control frames are derived from the BERs and the frame lengths.

Now we can drive the frame error rate, combining BER values for both header and payload as:

$$P_e = 1 - (1 - BER_{1Mbps})^{PHY} + (1 - BER_{11Mbps})^{(MAC+DATA)} \tag{30}$$

where, PHY is the length of the physical header, MAC is the length of the MAC header and DATA is the length of the packet payload.

B. Numerical results and discussions

The behavior of the transmission probability 'τ' is depicted in Fig.3 for basic access method as a function of SINR. The curves have been drawn for the capture threshold 6dB, number of contenting stations 10 and payload size 1024 bytes. The results shows that for increasing the channel quality, the transmission probability 'τ' increases and reaches the steady state, above which the channel is assumed as ideal. The transmission probabilities of our proposed model and model [17] are clearly highlighted in the above results. The Bianchi's transmission probability is depicted as horizontal lines due to independence of the Bianchi's model on both capture effects and channel errors.

Fig.4 shows the behavior of the saturation throughput as a function of the number of the contending stations for basic access mechanism. The curves have been drawn for the capture threshold 6dB, SINR 7dB and payload size 1024 bytes. The curves clearly show that, when the network load is moderate our proposed algorithm performs well. On the other hand, throughput can be higher than the model[17] for a low number of contending stations in the considered scenario. When the number of contending stations increases then the achievable throughput will approach the saturation throughput obtained in model[17].

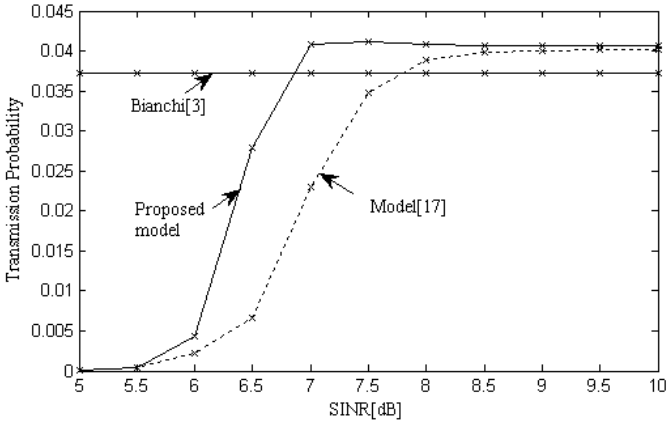


Figure 3. Transmission probability as a function of SINR for basic access mechanism. Curves have been obtained for the capture threshold 6dB, payload size 1024 bytes and number of contending stations 10.

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

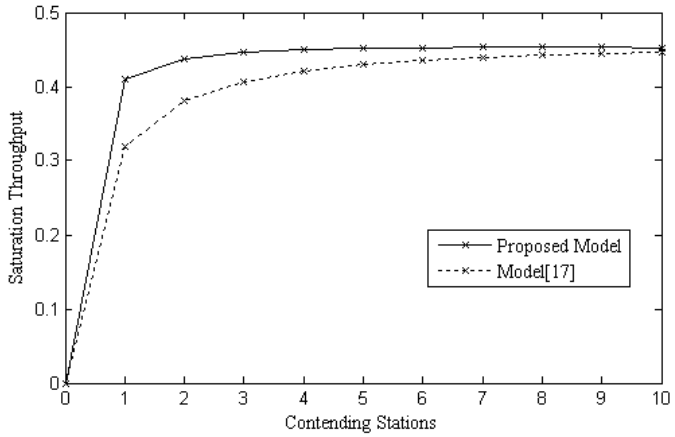


Figure 4. Saturation throughput for basic access mechanism as a function of the number of contending stations for capture threshold 6dB and SINR 7dB while payload size 1024 Bytes.

In order to assess throughput performance as a function capture threshold, Fig.5 shows throughput performance as a function SINR, for three different values of the capture threshold. Depending upon the channel quality as exemplified by the SINR on the abscissa in the figure, it could be preferable to operate at low capture threshold in order to gain higher throughput performance. The throughput predicted by Bianchi assuming SINR = ∞ and capture threshold = ∞, is depicted as a horizontal line along with the proposed model for comparison purpose. For decreasing capture threshold, the system throughput increases above the Bianchi's maximum achievable throughput performance. This is essentially due to the fact that, the capture effect tends to reduce the collision probability experienced by the contending stations which attempt simultaneous transmission.

Fig.6 shows the behavior of system throughput for basic access method as a function SINR, for two different payload sizes and for 5 transmitting stations. The upper curves are plotted for the payload size of 1024bytes and the bottom curves are plotted for the payload size of 128 bytes. In both curves, saturation throughput is depicted for two different values of capture threshold.

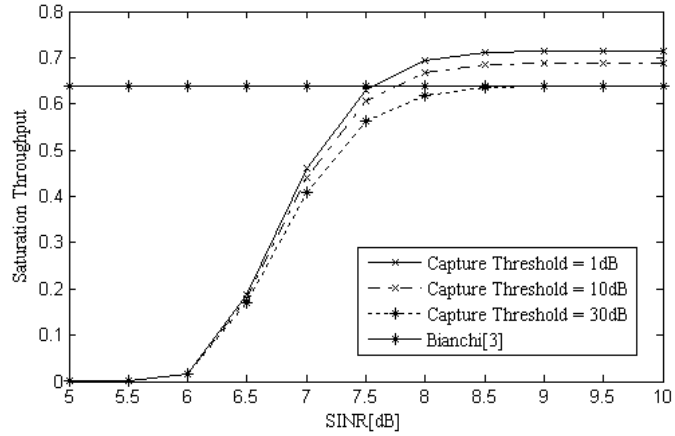


Figure 5. Saturation throughput for basic access mechanism as a function of SINR for capture thresholds 1dB, 10dB and 30dB, while payload size 1024 bytes and number of contending stations 5.

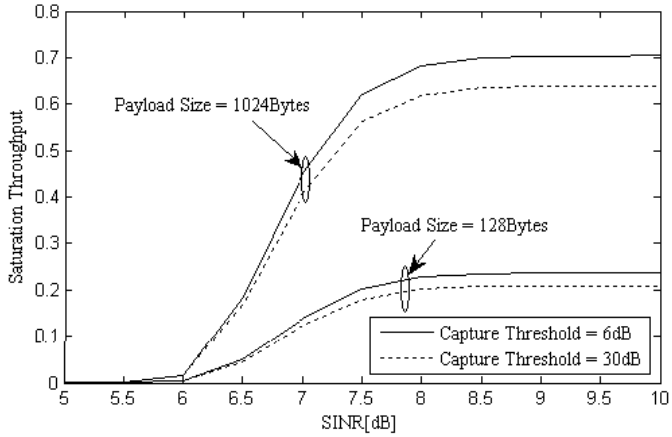


Figure 6. Saturation throughput for basic access mechanism as a function of SINR for payload sizes 1024 and 128 bytes, number of contending stations 5 and capture thresholds 6dB and 30dB.

Upon comparing the curves, it is easily seen that the system throughput performance is poor for low values of payload size. On the other hand, when the capture threshold is high, collision probability increases, that tend to reduce the throughput performance.

Fig.7 shows the behavior of saturation throughput for basic access method as a function SINR for two different capture thresholds. It can be easily noticed that, when channel errors are more, the achievable throughput is high due to proper rescheduling of contention window. On the other hand, for increasing capture threshold, throughput tends to reduce, as expected, in the presence of capture.

IV. CONCLUSION

In this paper we have proposed a new MAC protocol for IEEE802.11 Distributed Coordination Function taking into account of both erroneous channel and capture effects. This avoids unnecessary idle slots by differentiating noise lost packets from collision lost packets, increasing throughput considerably. It performs as well as IEEE 802.11 in noisy environment considering low traffic conditions. Using the

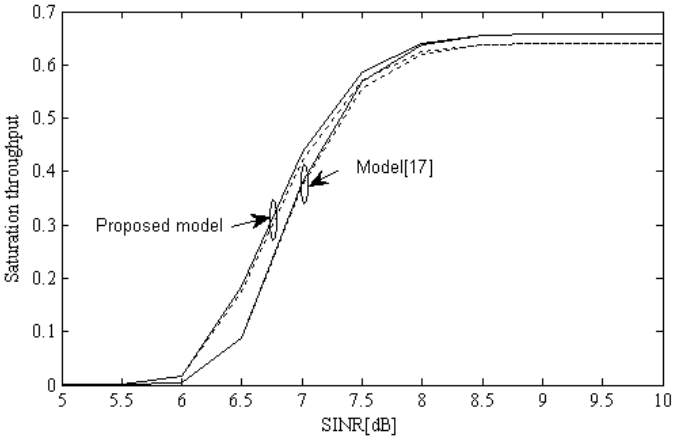


Figure 7. Saturation throughput for basic access mechanism as a function of SINR for capture thresholds 6dB and 24dB while payload size 1024 Bytes and number of contending stations 2.

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

proposed model we have evaluated the throughput performance of IEEE802.11 DCF for basic access method. Based on this model we derive a novel and generalized expression for the station's transmission probability, which is more realistic, such as non ideal channel conditions. To the best of our knowledge, this paper is the first to show the undesirable behavior of the standard backoff procedure when transmission losses occur, to develop a practical solution to this problem, and to give a theoretical performance analysis under homogeneous link conditions.

REFERENCES

[1] IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, P802.11, 1999.

[2] IEEE LAN/MAN Standards Committee. Part 15.2: Coexistence of wireless personal area networks with other wireless devices operating in unlicensed frequency bands, August 2003.

[3] G.Bianchi, " Performance Analysis of the IEEE 802.11 Distributed Coordination Function", IEEE Journal on Selected Area in Communications, vol.18, no.3, pp. 535-547, March 2000.

[4] F. Cali, M. Conti, E. Gregori, IEEE 802.11 protocol: design and performance evaluation of an adaptive backoff mechanism, IEEE J. Select. Area Commun. 18 (2000) 1774–1786.

[5] H. Wu, Y. Peng, K. Long, S. Cheng, J. Ma, Performance of reliable transport protocol over IEEE 802.11 wireless LAN: analysis and enhancement, in: INFOCOM 2002 – 21th Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, 2002, pp. 599–607.

[6] F. Cali, M. Conti, E. Gregori, "Dynamic tuning of the IEEE 802.11 protocol to achieve a theoretical throughput limit", IEEE/ACM Trans. Netw. Vol.8, no.6, pp.785-799, December 2000.

[7] Y.C. Tay, K.C. Chua, A capacity analysis for the IEEE 802.11 MAC protocol, Proceedings of Wireless Networks 7 (2001) 159–171.

[8] H.-T. Wu, Y.P., K. Long, S.-D. Cheng, J. Ma, Performance of reliable transport protocol over IEEE 802.11 wireless LAN: analysis and enhancement, in: Proceedings of IEEE INFOCOM 2002, New York, NY, June, 2002.

[9] Y. Xiao, Saturation performance metrics of the IEEE 802.11 MAC, in: Proceedings of the IEEE Vehicular Technology Conference (IEEE VTC 2003 Fall), Orlando, FL, USA, October 6–9, 2003.

[10] Z.-H. Velkov, B. Spasenovski, Saturation throughput—delay analysis of IEEE 802.11 DCF in fading channel, in: Proceedings of the IEEE ICC_03, Anchorage, AK, USA, May, 2003.

[11] Xiaohui Xu, Xiaokang Lin, "Throughput Enhancement of the IEEE 802.11 DCF in fading channel", In proc. IEEE international conference on wireless and optical communications networks, Augest 2006, Bangalore, India.

[12] J. Yin, X. Wang, and D. P. Agrawal, "Impact of Bursty Error Rates on the Performance of Wireless Local Area Network (WLAN)", Ad Hoc Networks, vol. 4. no.5, pp. 651-668, 2006.

[13] Y.Kwon, Y.Fang, and H.Latchman, "Design of MAC protocols with fast collision for wireless local area networks," IEEE Trans. Wireless commun., vol. 3, no. 3, pp. 793-807, May 2004.

[14] Hadi Minooei, Hassan Nojumi "Performance evaluation of a new backoff method for IEEE 802.11", Elsevier journal on Computer Communication, vol. 30, issue. 18 pp. 3698-3704, December 2007.

[15] David Malone, Ken Duffy, and Doug Leith, "Modeling the 802.11 Distributed Coordination Function in Nonsaturated Heterogeneous Conditions", IEEE/ACM Trans. Networking, vol.15, no.1, pp 159-172, February 2007.

[16] Ergen and P. Varaiya, Formulation of Distributed Coordination Function of IEEE802.11 for Asynchronous Networks: Mixed Data Rate and Packet Size, IEEE Transaction on Vehicular Tech, Vol.57, no.1, pp.436-447, January 2008.

[17] F. Daneshgaran, M. Laddomada, F. Mesiti, and M. Mondin, "Saturation Throughput Analysis of IEEE 802.11 in Presence of Ideal Transmission Channel and Capture Effects", IEEE Trans. Commun., vol. 56, no. 7, pp. 1178-1188, July 2008.

[18] P.Kumar, A.Krishnan, K.Poongodi and T.D.Senthilkumar, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function in High Interference Wireless Local Area Networks considering Capture Effects", In Proc. of IEEE International Advanced Computing Conference(IACC), Vol.1, pp. 234-456, March 2009.

[19] T.D. Senthilkumar, A. Krishnan, P. Kumar, Nonsaturation Throughput Analysis of IEEE 802.11 Distributed Coordination Function. In *Proc. of IEEE International Conference on Control Communication and Automation (Indicon'08),* IIT Kanpur, India, pp. 154-158, 2008.

[20] Z. Hadzi-Velkov and B. Spasenovski, "Capture effect in IEEE 802.11 basic service area under influence of Rayleigh fading and near/far effect", In Proc. of 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Vol.1, pp.172 - 176, Sept. 2002.

[21] J. G. Proakis, Digital Communications, New York: Mc-Graw Hill, 2001 (4th ed).

[22] Y. S. Liaw, A. Dadej, and A. Jayasuriya, "Performance analysis of IEEE 802.11 DCF under limited load," in *Proceedings of Asia-Pacific Conference on Communications*, pp. 759–763, Perth, Western Australia, October 2005.

[23] T.D. Senthilkumar, A. Krishnan, P. Kumar, New Approach for Throughput Analysis of IEEE 802.11 in AdHoc Networks. In *Proc.of IEEE International Conference on Control Communication and Automation (Indicon'08) ,* IIT Kanpur, India, pp. 148-153, 2008.

[24] Q. Pang, S.C. Liew, and V. Leung, "Design of an effective loss-distinguishable MAC protocol for 802.11 WLAN," *IEEE Communications Letters*, vol. 9, no. 9, pp.781-783, Sep. 2005.

[25] P. Chatzimisios, A.C. Boucouvalas, and V. Vitsas, "Performance analysis of IEEE 802.11 DCF in presence of transmission errors," *IEEE ICC 2004*, pp. 3854-3858, Jun. 2004.

[26] J. Yeo and A. Agrawala, "Packet error model for the IEEE 802.11 MAC protocol," *IEEE PIMRC 2003*, pp.1722-1726, Jan. 2003.

[27] J. Yin, X. Wang, and D.P. Agrawal, "Optimal packet size in error-prone channel for IEEE 802.11 distributed coordination function," *IEEE WCNC 2004*, pp.1654-1659, Mar. 2004.

AUTHORS PROFILE

Ponnusamy Kumar received his B.E degree from Madras University, Chennai, India, in 1998, and the M.Tech degree from Sastra University, Tanjavore, India, in 2002. Since 2002 he is working as a Assistant Professor in K.S.Rangasamy College of Technology, Tamilnadu, India. He is currently pursuing his Ph.D degree under Anna University, Chennai, India. His research is in the general area of wireless communication with emphasis on adaptive protocols for packet radio networks, and mobile wireless communication systems and networks. Mr.P.Kumar is a member in IETE and ISTE.

A. Krishnan received his Ph.D degree from IIT Kanpur, Kanpur, India. He is currently a professor with K.S.Rangasamy College of Technology, Tiruchengode, Tamilnadu, India. He has published over 150 papers in national and international journals and conferences. His research interests are in the area of communication networks, transportation, and hybrid systems. Dr. A. Krishnan is a senior member in IEEE, and member in IETE and ISTE.

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

Performance Evaluation of Unicast and Broadcast Mobile Ad-hoc Networks Routing Protocols

Sumon Kumar Debnath[1], Foez Ahmed[2], and Nayeema Islam[3]

[1]Dept. of Computer Science and Telecommunication Engineering, Noakhali Science and Technology University, Bangladesh
[2]Dept. of Computer Networks Engineering, College of Computer Science, King Khalid University, Kingdom of Saudi Arabia
[3]Department of Information & Communication Engineering, Rajshahi University, Rajshahi- 6205, Bangladesh

***Abstract*— Efficient routing mechanism is a challenging issue for group oriented computing in Mobile Ad Hoc Networks (MANETs). The ability of MANETs to support adequate Quality of Service (QoS) for group communication is limited by the ability of the underlying ad-hoc routing protocols to provide consistent behavior despite the dynamic properties of mobile computing devices. In MANET QoS requirements can be quantified in terms of Packet Delivery Ratio (PDR), Data Latency, Packet Loss Probability, Routing Overhead, Medium Access Control (MAC) Overhead and Data Throughput etc. This paper presents an in-depth study of one-to-many and many-to-many communications in MANETs and provides a comparative performance evaluation of unicast and broadcast routing protocols. Dynamic Source Routing protocol (DSR) is used as unicast protocol and BCAST is used to represent broadcast protocol. The performance differentials are analyzed using ns-2 network simulator varying multicast group size (number of data senders and data receivers). Both protocols are simulated with identical traffic loads and mobility models. Simulation result shows that BCAST performs better than DSR in most cases.**

Keywords-MANETs, DSR, BCAST, Unicast, Broadcast, Random way point model

I. INTRODUCTION

A mobile ad hoc network is a self-organizing network comprising wireless mobile nodes that move around arbitrarily and can able to communicate among themselves using wireless radios, without the aid of any preexisting infrastructure [1]. Each participating mobile node can act as sender, receiver and even as a router at the same time and able to build, operate and maintain these networks [2].

Due to limited radio coverage of these wireless devices efficient support of group oriented communication is extremely critical in most MANET applications. In MANET group communications issues differ from those in wired environments for the following reasons: The wireless medium has variable and unpredictable characteristics. The signal strength and propagation fluctuate with respect to time and environment resulting disconnection of the network at any time even during the data transmission period [3]. The strength of the received signal depends on the power of the transmitted signal, the antenna gain at the sender and receiver, the distance between two mobile nodes, the obstacles between them, and the number of different propagation paths the signals travel due to reflection. Further node mobility also creates a continuously changing communication topology in which existing routing paths break and new ones form dynamically. Since MANETs have limited channel bandwidth availability and low battery power, their algorithms and protocols must conserve both bandwidth and energy [3]. Wireless devices usually use computing components such as processors, memory, and I/O devices, which have low capacity and limited processing power. Thus their communication protocols should have lightweight computational and information storage capability fulfilling some key features like robustness, simplicity and energy conserving. For this reason, several prominent unicast, multicast and broadcast protocols deployed in static wired networks that can not perform well in ad-hoc networks [3, 4].

In-group oriented communication system, routing protocols can be classified into three main categories [5, 6, 7] based on the number of senders and receivers in MANETs. Unicast communication is the point-to-point transmission with one sender and one receiver. While unicasting is a simple mechanism for one-to-one communication, for one-to-many or many-to-many dissemination it brings the network to its knees due to huge bandwidth demands [8]. This can also introduce significant traffic overhead, sender and router processing, power consumption, high packet latency and poor throughput in the network. To minimize these overhead for one-to-many or many-to-many communication multicast and broadcast Ad-Hoc routing protocols play an important role. Multicast communications are both one-to-many and many-to-many traffic pattern [9] i.e. to transmit a single message to a select group of recipients where as in broadcast routing communications is one-to-all traffic pattern. It is a basic mode of operation in wireless medium that provides important control and route establishment functionality for a number of unicast and multicast protocols. When designing broadcast

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

protocols for ad hoc networks, developers seek to reduce the overhead such as collision and retransmission or redundant retransmission, while reaching all the network's nodes. In practice, the scope of the broadcast is limited to a broadcast domain. Broadcasting is largely confined to local area network (LAN) technologies, most notably Ethernet and Token Ring, where the performance impact of broadcasting is not as large as it would be in a wide area network. Because Broadcasting is used to carry huge amount of traffic and requires more bandwidth, neither X.25 nor frame relay supply a broadcast capability, nor Internet explicitly support broadcasting at the global level [10].

This paper compares two Ad Hoc routing protocols: unicast reactive DSR and BCAST protocol over group oriented communication system. Performance comparisons are based on Shadowing path loss model and Random way point mobility model. The simulation of two routing protocols focuses on their differences in their dynamic behaviors that can lead to performance differences.

II. UNICAST AND BROADCAST ROUTING PROTOCOLS

A. Dynamic Source Routing(DSR) Protocol

DSR [11, 12] is an on-demand unicast reactive source-routed routing protocol. This means the source node determines the complete sequence of route information between source and destination and explicitly lists each hop of the path in the packet's header. Route is determined dynamically without any prior configuration necessary. The intermediate nodes do not require huge memory resources because they do not need to maintain consistent global routing information in order to route the packets that they forward. The basic operation of DSR contains two phases: route discovery and route maintenance. Route Discovery mechanism is used to find a source route to destination only when source attempts to send a packet to destination and does not already knows a route. To reduce the cost of Route Discovery, each node maintains a Route Cache of source routes it has learned or overheard. Route Maintenance is the mechanism used to detect if the network topology has changed such that it can no longer use its route to the destination because some of the nodes listed on the route have moved out of range of each other.

B. Broadcast Routing (BCAST) Protocol

BCAST is an optimized scalable broadcast routing protocol [13]. It keeps track of one-hop and two-hop neighbor knowledge information that are exchanged by periodic "Hello" messages. Each "Hello" message contains the node's IP address and list of known neighbors. When a node receives a "Hello" packet from all its neighbors, it has two-hop topology information i.e. only packets that would reach additional neighbors are re-broadcast. For example if a node, *B* receives a broadcast packet from another node *A*, it knows all neighbors of *A*. If *B* has neighbors not covered by *A*, it issues the broadcast packet with a random delay. During this delay, if *B* receives another copy of this broadcast from *C*, it can check whether its own broadcast will still reach new neighbors. If this is no longer the case, it will drop the packet. Otherwise, the process continues until *B*'s timer goes off and *B* itself rebroadcasts the packet.

The determination of Random delay time is very critical. To solve this problem a dynamic strategy is suggested in literature [13]. Each node searches its neighbor table for the maximum neighbor degree of any neighbor node, say *MAX*. If its own node degree is N, it calculates the random delay as *MAX*/*N*. Every node also buffers the most recent *X* packets. *X* can be any arbitrary integer number. To keep the memory requirement at each node low; set *X* to a small number. This mechanism improves the packet delivery ratio in BCAST.

When a node receives a packet with sequence number *N* from source node *A*, it checks whether it also received packet *N*-1 from the same source. If not, it issues a one-hop broadcast to the neighbors, asking for retransmission of this packet by sending Negative Acknowledgement, NACK(N-1, A) message. Each neighbor, upon receiving the NACK packet, will check its local buffer and if they have this packet buffered, will schedule a retransmission. To reduce collisions, the NACKs and the packet retransmissions are jittered randomly by few milliseconds.

III. SIMULATION MODEL

This section describes the simulation tools and parameters chosen to simulate the routing protocols

A. Simulation Environment

Network Simulator NS-2 [14, 15] is chosen to compare DSR and BCAST routing protocols. NS-2 is discrete event packet-level simulators with CMU's Monarch group's mobility extensions. It includes implementations of models of signal strength, radio propagation, wireless medium contention, capture effect, and node mobility. A simulation model with MAC and physical models are used to study the interlayer interaction and their performance. An unslotted carrier sense multiple access (CSMA) technique with collision avoidance (CSMA/CA) is used to transmit the data packets. In this experiment, the Distributed Coordination Function (DCF) of IEEE 802.11 for wireless LAN is used as MAC layer. The simulated radio interface model is the Lucent WaveLAN. WaveLAN is modeled as shared-media radio with channel capacity of 2Mbits/sec and transmission range of 250m and the carrier sensing range is 471.5m. All protocols use an interface queue (IFQ) of 50 packets. The IFQ is a FIFO priority queue where routing packets gets higher priority than data packets. All MAC and Network layer operations of the wireless network interfaces are logged in trace files.

B. Radio Propagation Model

The shadowing path loss model [15] is used in this simulation study. It attempts more realistic situation than free space and two-ray path loss models. It takes into account multi-

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

path propagation effects. Both free space and two-ray models predict the mean received signal strength as a deterministic function of distance and consequently represent communication radius as an ideal circle. But in realistic environment, when the fading effects are considered it can be seen that, the received power at a certain distance is a random variable. Hence shadowing model is widely used in real environment.

The available parameters that are used in our simulation code are shown in table I.

TABLE I. PARAMETERS USED IN SIMULATION

Parameters	Value	Comment
Transmission Range	250m	Fixed (Considered)
Frequency	$914 \times 10^{6} Hz$	Fixed (Considered)
Path Loss Exponent	2.0	Fixed (Considered)
Standard Deviation	4.0	Fixed (Considered)
Reference Distance	1.0m	Fixed (Considered)
CPThreshold	10.0 Watt	Fixed (Considered)
RXThreshold	6.76252×10^{-10}	Calculated
CSThrehold	2.88759×10^{-11}	Calculated (RXThreshold*0.0427)
Power (Pt)	0.28183815 Watt	Fixed (Considered)
System Loss	1.0	Fixed (Considered)

C. Traffic and Mobility Model

In this simulation Continuous bit rate (CBR) traffic sources are used. The source-destination pairs are spread randomly over the network. Only 512- byte data packets are used. The number of source destination pairs and the packet-sending rate in each pair is varied to change the offered load in the network.

A mobility model accurately represents the movement of mobile nodes in MANET. Random waypoint mobility model [16] *(*RWM) is used in this simulation study. The model includes networks with 50 mobile nodes placed on a site with dimensions 1500×300 meters. Each packet starts its journey from a random location to a random destination with a randomly chosen speed (uniformly distributed between 0–20 m/s). Once the destination is reached, another random destination is targeted after a pause time and then repeats the process. The pause time, which affects the relative speeds of the mobiles, is also varied. Five randomly generated scenarios are run for each parameter combination, and each point in the graphs is the average results of these five scenarios. Identical mobility and traffic scenarios are also used across protocols to gather fair results. In RWM model, *Pause Time* and *Max Speed* of a mobile are the two key parameters that determine the mobility behavior of nodes. If the node movement is small and the *Pause Time* is long, the topology of Ad Hoc network becomes relatively stable. On the other hand, if the node moves fast and the pause time is small; the topology is expected to be highly dynamic.

IV. PERFORMANCES METRICES

The performance of DSR and BCAST protocols are compared using the following important Quality of Service (QoS) metrics

Packet Delivery Ratio (PDR): The ratio of the number of packets received by the CBR sinks at the final destination to those generated by the CBR sources.

Packet Latency: This includes all possible delays caused by buffering during route discovery, queuing delay at the interface queue, retransmission delays at the MAC, propagation and transfer times [17]. The lower the packet latency the better the application performance as the average end-to-end delay is small.

Normalized routing Load (NRL): The ratio of the number of routing packets sent to the number of data packets received. Each hop-wise transmission of these packets is counted as one transmission [18].

Normalized MAC Load (NML): The number of routing, Address Resolution Protocol (ARP), and control packets (e.g., RTS, CTS and ACK) transmitted by the MAC layer, including IP/MAC headers for each delivered data packet [18]. It considers both routing overhead and the MAC control overhead. This metric also accounts for transmission at each hop.

Throughput: The ratio of the total data received by the end user and the connection time [19]. A higher throughput directly impacts the user's perception of the QoS.

V. SIMULATION RESULTS AND DISCUSSIONS

The results of this simulation study are separately considered into two sections.

- Varying Number of data Senders:
- Varying Number of data Receivers

A. Effect of number of senders on QoS metrics

To investigate the effect of number of senders on the performance of DSR and BCAST, the data send rate and number of data receivers is kept constant at 2 packets/sec. and 20 respectively. The numbers of data senders are increased

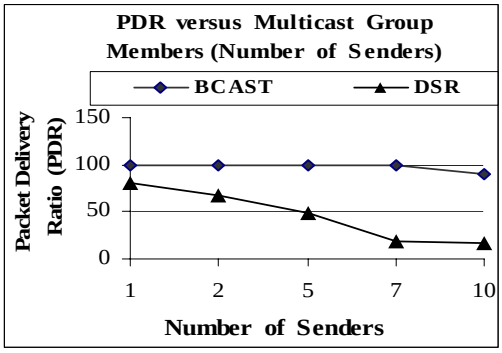
http://sites.google.com/site/ijcsis/
ISSN 1947-5500

from 1 to 10 and several QoS metrics are measured and plotted into logarithmic graphs. For the fairness of protocols comparison and network performance, each ad hoc routing protocol is run over the same set of scenarios. Table II. Shows Simulation parameters for the different Senders Scenarios

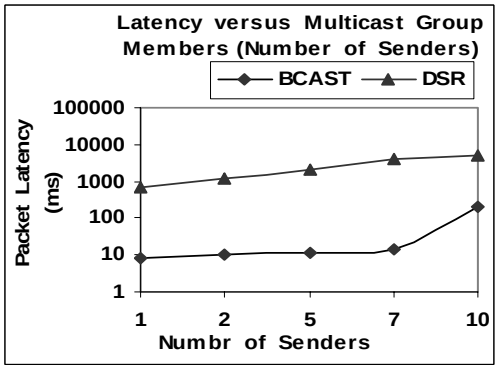
TABLE II. SIMULATION PARAMETERS FOR THE DIFFERENT SENDERS SCENARIOS

Parameter	Value
Number of senders (variable)	1, 2, 5, 7 and 10
Number of receivers (keep constant)	20
Pause Time	0 m/s
Max. Speed	20 m/s
Antenna Range	250 m
CBR Rate	2 packets/sec.
Simulation Time	200 s

The PDR and data packet latency simulation results as a function of number of Senders are given in fig 1.



(a)



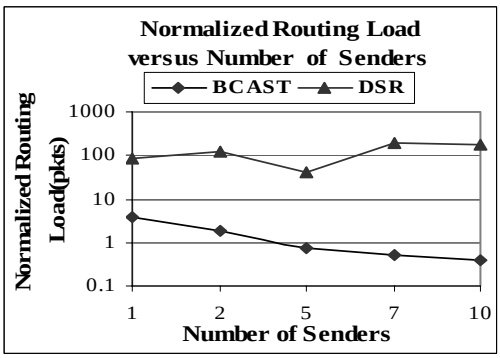
(b)

Figure 1. PDR (a) and Latency (b) results as a function of Number of Senders.

From fig 1(a) it is observed that with increasing number of senders, the PDR of BCAST protocol is higher and relatively consistent as compare to DSR routing protocol. A relatively high PDR is a desirable property for the scalability of Ad Hoc routing protocols. Unicast DSR shows lower PDR than BCAST for one-to-many and many-to-many communication. This is due to the fact that with relatively many senders (two or more) the traffic sources are more spread throughout the MANET and hence the overall performances of on-demand unicast DSR deteriorate rapidly. For example in DSR with five senders and 30 receivers, each sender needs to generate and maintain 30 simultaneous unicast connections to connect 30 multicast group members. Hence overall 150 unicast connections are required. Each data source also requires generating 30 data packets. This provides an inefficient use of wireless medium and causes congestion in the network. Since for many senders the PDR value drops drastically, the performance of DSR protocol is not very attractive for many-to-many application and it also increases the packet transmission cost. BCAST shows higher packet delivery ratio for most scenarios. Fig 1(a) also shows that for five senders, the PDR of BCAST and DSR are 99.45% and 47.82% respectively. This is because that BCAST has less redundancy and dynamically selects only a subset of nodes to re-broadcast a packet. It keeps 2-hop neighbor topology information and each node also buffers most recent few packets. A NACK based retransmission scheme of BCAST protocol further increase PDR.

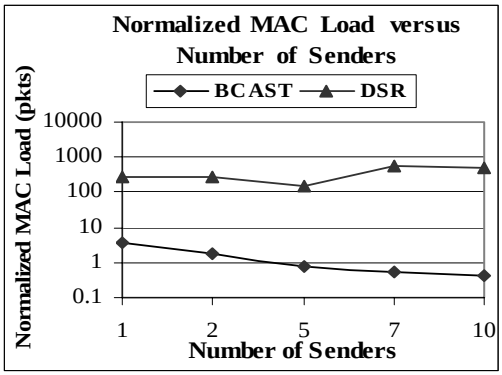
From fig 1(b) it is observed that the average packet latency increases with increasing number of senders. BCAST protocol performs better than DSR in this case. For example for seven senders, the packet latency of BCAST and DSR are 14.43 ms and 3.86 μs respectively. This is due to the fact that DSR maintains unicast connections. As the number of sender's increases, it requires to generate more packets in order to reach the group members, more routing packets causes delay in the interface queue before reaching the intended destination. This causes more packet delay with increasing the number of senders. Since BCAST maintains broadcast connections and keep two hops topology information, the average packet delay is significantly lower than DSR. Lower packet latency is the desirable property for real-time applications because these applications can tolerate loss but very sensitive to delay. Hence BCAST is more effective for real time applications.

The NRL and NML simulation results as a function of number of Senders are given in fig 2.



(a)

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



(b)

Figure 2. .NRL (a) and NML (b) results as a function of Number of Senders

From fig 2(a) it can be observed that the NRL of DSR is higher than BCAST and hence provides poor performance. This is because that with increasing the number of senders DSR requires more routing packets to maintain unicast connections among group members. For example for five senders, NRL of DSR and BCAST are 44.23 packets and 0.735 packets respectively. From fig 2(b), it is also observed that in DSR normalized MAC load is also extremely high than BCAST protocols. In this case almost all MAC transmissions are unicast, a high fraction these transmitted packets are MAC layer control packets (RTS, CTS and ACK). Due to unicast nature, as the number of senders increase DSR requires to generate more MAC control packets. BCAST gives better performance in this case. Since in BCAST all MAC transmissions are multicast, it generates only fraction of MAC layer control packets than DSR. This effect results lower transmission collision and offers high packet delivery guarantee. For example for 10 senders the NML of BCAST and DSR protocols are 0.41 and 511.03 packets respectively.

The data throughput results as a function of number of Senders are given in fig 3.

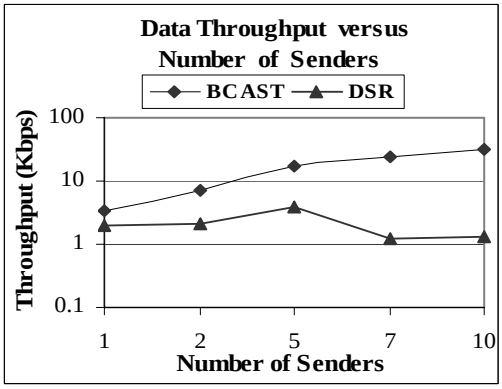


Figure 3. .Data throughput results as a function of Number of Senders.

This throughput result is consistent with the control overhead (NRL and NML) shown in fig 3. This is because that the decay of overhead is directly related to the throughput growth. From fig 3 it can also be observed that with increasing the number of senders, the data throughput increases and the throughput of DSR does not increases too drastically and provides poor performance. For few senders (one to five) DSR throughput increases but lower than BCAST protocols. With more senders the DSR throughput highly degrades. Since DSR packet drop probability increases with increasing the number of senders, it affects the DSR throughput. It can be observed that as the number of senders increase, the throughput of BCAST increases and the maximum throughput are achieved for BCAST protocol. For example at five senders, fig 3 mentions that the data throughputs of BCAST and DSR protocols are 17.54 kbps and 3.91 kbps respectively.

B. Effect of number of receivers on QoS metrics

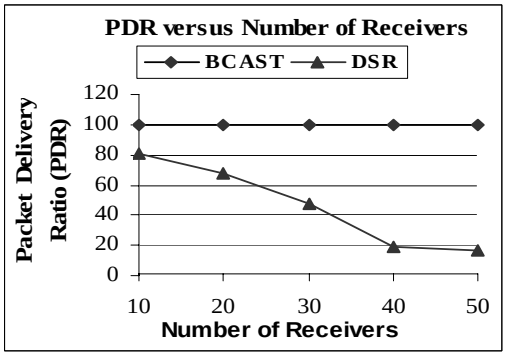
In this case, to evaluate the performance of DSR and BCAST, the data send rate and number of data senders are kept constant at 2packets/sec. and 05 respectively. The numbers of multicast receivers are varied and QoS metrics are measured.

The simulation parameters considered for performance evaluation are provided in table III.

TABLE III. SIMULATION PARAMETERS FOR THE DIFFERENT RECEIVERS SCENARIOS

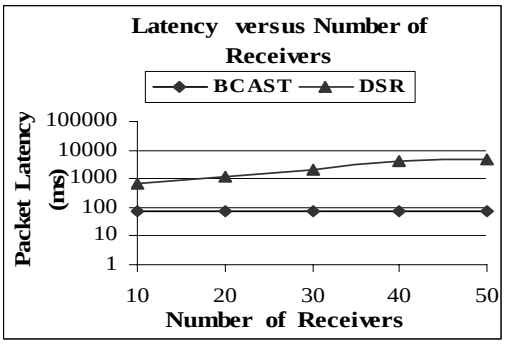
Parameter	Value
Number of senders (keep constant)	05
Number of receivers (variable)	10, 20, 30, 40 and 50
Pause Time	0 m/s
Max. Speed	20 m/s
Antenna Range	250 m
CBR Rate	2 packets/sec.
Simulation Time	200 s

The PDR, average packet latency and throughput results as a function of number of receives is shown in fig 4.

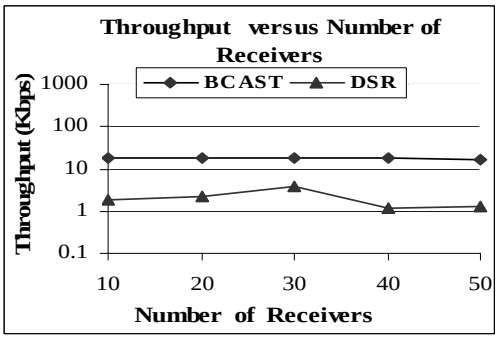


(a)

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



(b)

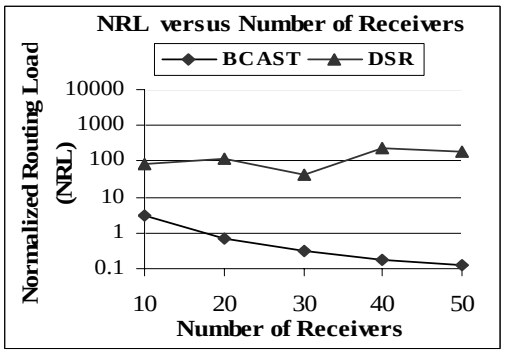


(c)

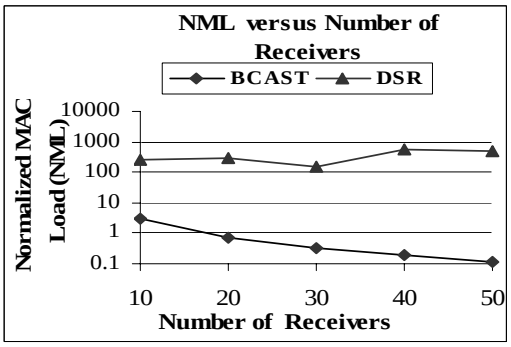
Figure 4. .PDR (a), Latency (b) and Data throughput (c) results as a function of Number of Receivers.

From fig 4(a) and fig 4(b), it is observed that, DSR gives similar PDR and Latency results as in the case of number of senders. BCAST protocol also shows better performance and these performances are independent of the number of multicast receivers. From fig 4(c), it is observed that, the throughput performance of BCAST is independent of the number of multicast receivers. BCAST protocol shows better performance in this case. In DSR when the number of multicast receivers below 30, DSR throughput performance increases with increasing the number of receivers. After this the throughput deteriorates again.

The NRL and NML simulation results as a function of number of multicast receivers are given in fig 5.



(a)



(b).

Figure 5. .NRL (a) and NML (b) results as a function of versus Number of Receivers

From fig 5(a) and fig 5(b), it is observed that the normalized routing and MAC load of DSR and BCAST protocols are also similar as in the case of number of senders.

VI. CONCLUSIONS

There are number of alternatives when delivering data from one or more senders to a group of receivers such as setting up dedicated unicast connections from each sender to each receiver, employing a unicast, multicast protocol and broadcasting packet to every node. This paper compares the performance of BCAST and DSR routing protocols over group communication in MANETs. BCAST is an optimized neighbor knowledge based broadcast protocol provides robust performance with less delay time (keeps two hop neighbor information and minimizes network congestion) and less traffic overhead (partial source route) in terms of PDR, latency, normalized routing and MAC load, packet drop probability and data throughput. On-demand unicast DSR protocol suffers more from scalability issue as the number of data senders and data receivers increase. For scenarios with N senders and M receivers, $N \times M$ unicast connections have to be discovered and maintained by the underlying unicast routing algorithm. This introduces substantial overhead and causes high network load. Unicast DSR is also more sensitive to data send rate. The simulation result shows that the broadcast protocol BCAST work very well in most scenarios and are more robust even with high traffic environments.

REFERENCES

[1] G. Holland and N. Vaidya, (2002), "Analysis of TCP Performance over Mobile Ad Hoc Networks," 5th Annual Int'l. Conf. Mobile Comp. And Net, pp. 275–288, 2002 [Academic Publishers. Manufactured in The Netherlands].

[2] E. Baburaj and V. Vasudevan (2008), "An Intelligent Multicast Ad-hoc On demand Distance Vector Protocol for MANETs" Journal of Networks, Vol. 3, No. 6, June 2008.

[3] P. Mohapatra, C. Gui, and J. Li, (2004), "Group communications in mobile ad hoc networks**",** University of California, Davis, Computer Volume 37, Issue 2, Pages 52 – 59, Feb 2004 [published by the IEEE Computer Society].

[4] G. H. Lynn, (2003), "ROMR: Robust Multicast Routing in Mobile Ad-hoc Networks", Ph.D. Thesis, University of Pittsburgh, November 25, 2003.

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

[5] L. M. Feeney, (1999), "A Taxonomy for Routing Protocols in Mobile Ad Hoc Networks", Swedish Institute of Computer Science, sept 29, 1999.

[6] X. Hong, K. Xu and M. Gerla, (2002), "Scalable Routing Protocols for Mobile Ad Hoc Networks", University of California, Los Angeles CA 90095, IEEE network, 2002.

[7] P. Kuosmanen, F. D. Forces and N. Academy, (2002), "Classification of Ad Hoc Routing Protocols", 2002.

[8] A. S. Tanenbaum, (2003), "Computer Networks", Fourth Edition, 2003.

[9] C. E. Perkins, E. Royer, S. R. Das, and M. K. Marina, (2001), "Performance Comparison of Two On-demand Routing Protocols for Ad hoc Networks", IEEE Personal Communications Magazine special issue on Ad hoc Networking, pp. 16-28, February 2001.

[10] B. A. Forouzan, (2000), "TCP/IP Protocol Suite", Third Edition, Tata McGraw-Hill Edition, 2000.

[11] D. B. Johnson and D. A. Maltz, (2004), "The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR)", IETF MANET working group Internet draft, 19 July 2004.

[12] X. Li, (2006), "Multipath Routing and QoS Provisioning in Mobile Ad hoc Networks", Ph.D. Thesis, Department of Electronic Engineering Queen Mary, University of London, April, 2006.

[13] T. Kunz, (2003), "Reliable Multicasting in MANETs", Contract Report, Defence R&D Canada – Ottawa Communications Research Centre, July 2003.

[14] "The Network Simulator ns-2", http://www.isi.edu/nsnam/ns/.

[15] K. Fall, and K. Varadhan, (2008), "The ns Manual (formerly ns Notes and Documentation)", The VINT Project, UC Berkeley, LBL, USC/ISI, and Xerox PARC, September 2008.

[16] J. Broch, D. A. Maltz, D. B. Johnson, Y-C. Hu and J. Jetcheva, (1998), "A Performance Comparison of Multihop Wireless Ad Hoc Network Routing Protocols", Proc. IEEE/ACM Mobicom Conference, pp.85-97, zoctober 1998.

[17] M. K. Marina and S. R. Das, (2003), "Ad hoc on-demand multipath distance vector routing", Technical Report, Computer Science Department, Stony Brook University, April 2003

[18] G. Jayakumar and G. Ganapathy, (2007), "Performance Comparison of Mobile Ad-hoc Network Routing Protocol", IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.11, November 2007.

[19] L. Wang, L., Y. Shu, M. Dong, L. Zhang and O. Yang, (2001), "Adaptive Multipath Source Routing in Ad Hoc Networks", IEEE International Conference on Communications, ICC 2001.

AUTHORS PROFILE

Sumon Kumar Debnath obtained his M.Sc degree in Information & Communication Engineering from Rajshahi University, Bangladesh during the year 2009. He has joined Noakhali Science & Technology University, Bangladesh and working as a lecturer in the dept. of Computer Science and Telecommunication Engineering. He is also working as an Assistant Proctor in that university. His research interests include Mobile Ad-hoc networks, Vehicular Ad-hoc networks, Sensor Network, MIMO, OFDM, and MCCDMA.

Foez Ahmed joined as a Lecturer in the Department of Electronics and Communication Engineering, Northern University Bangladesh (NUB), Dhaka, Bangladesh in the year of 2008. At present he is on leave from Northern University and working as a Lecturer with the Dept of Networks and Communication Engineering, College of Computer Science, King Khalid University, Kingdom of Saudi Arabia. He did his under graduation and post graduation in Information and Communication Engineering in 2007 and 2009 respectively from Rajshahi University, Bangladesh. He has received various Awards and Scholarships for the under graduation and post graduation results. His research interests include Mobile Ad-hoc Networks and Routings, Cognitive Radio Networks, Cooperative Communications, Sensor Networks and Sparse Signal Processing in Wireless Communication.

Nayeema Islam was born 1976 in Bangladesh. She received her M.S. in Telecommunication Engineering from Asian Institute of Technology, Thailand in 2004. Also she received her M.Sc. and B.Sc. in Computer Science and Technology from Rajshahi University, Bangladesh in 1998 and 1997 respectively. She is presently working as Assistant Professor in the department of Information and Communication Engineering, University of Rajshahi since 2004. Her fields of interest include Telecommunications, computer networking, mobile ad-hoc networks and QoS routing.

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

Deriving Relationship Between Semantic Models - An Approach for cCSP

Shamim H. Ripon [1], Michael Butler [2]

[1] *Department of Computing Science, University of Glasgow, UK*

.

[2] *School of Electronics and Computer Science, University of Southampton, UK*

***Abstract*—Formal semantics offers a complete and rigorous definition of a language. It is important to define different semantic models for a language and different models serve different purposes. Building equivalence between different semantic models of a language strengthen its formal foundation. This paper shows the derivation of denotational semantics from operational semantics of the language cCSP. The aim is to show the correspondence between operational and trace semantics. We extract traces from operational rules and use induction over traces to show the correspondence between the two semantics of cCSP.**

***Index Terms*—Compensating CSP, semantic relationship, trace semantics, operational semantics.**

I. Introduction

A formal semantics offers a complete, and rigorous definition of a language. Operational and denotational semantics are two well-known methods of assigning meaning to programming languages and both semantics are necessary for a complete description of the language. Denotational semantics associates an element of a semantic domain to each expression in the language and the semantic is compositional. Traces are one of the ways to define denotational semantics. A trace gives the global picture of the behaviour. The common way of defining operational semantics is to provide state transition systems for the language, where the transition system models the computation steps of expressions in the language and allows the formal analysis of the language.

Compensating CSP (cSCP) [1] is a language defined to model long running business transactions within the framework of Hoare's CSP [2] process algebra. Business transactions need to deal with faults that can arise at any stage of the transactions. *Compensation* is defined in [3] as an action taken to recover from error in business transactions or cope with a change of plan. cCSP provides constructs for orchestration of compensations to model business transactions. With the introduction of the language, both traces [1] and operational [4] semantics have been defined. Both semantics have valuable non-overlapping application and we want to use them both. The key question is *"How they are related?"*.

This paper draws the correspondence of two different semantic representation of a language which strengthen the formal foundation of the language. In particular, the aim is to accomplish the unification between operational and denotational approach of cCSP. The unification is based on the approach where we use the transition rules from operational semantics to derive the traces and then show that these derived traces correspond to the original traces by using induction over the derived traces. Completing the derivation means that any of the presentations can be accepted as a primary definition of the meaning of the language and each of the definitions can even safely and consistently be used at different times and for different purposes.

The reset of the paper is organised as follows. A brief overview of cCSP along with an example is given in Section II. The trace and the operational semantics of cCSP are outlined in Section III. We describe the how we define and prove a relationship between the semantic models in Section IV. We define theorems and supporting lemmas to establish the relationship for both standard and compensable processes. We outline some lessons from the experiment and then summarise some related work in Section V and Section VI respectively. We draw our conclusion in Section VII.

II. Compensating CSP

The introduction of the cCSP language was inspired by two ideas: transaction processing features, and process algebra. Like standard CSP, processes in cCSP are modelled in terms of the atomic events they can engage in. The language provides operators that support sequencing, choice, parallel composition of processes. In order to support failed transaction, compensation operators are introduced. The processes are categorised into *standard*, and *compensable* processes. A standard process does not have any compensation, but compensation is part of a compensable process that is used to compensate a failed transaction. We use notations, such as, $P, Q, ..$ to identify standard processes, and $PP, QQ, ..$ to identify compensable processes. A subset of the original cCSP is considered in this paper, which includes most of the operators, is summarised in Fig. 1.

The basic unit of the standard processes is an atomic event (A). The other operators are the sequential (P ; Q), and the parallel composition ($P \parallel Q$), the choice operator ($P \Box Q$), the interrupt handler ($P \rhd Q$), the empty process *SKIP*, raising an interrupt *THROW*, and yielding to an interrupt *YIELD*. A process that is ready to terminate is also willing to yield to an interrupt. In a parallel composition, throwing an interrupt by one process synchronises with yielding in another process. Yield

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

Standard Processes:		**Compensable Processes:**	
$P, Q ::= A$	(atomic event)	$PP, QQ ::= P \div Q$	(compensation pair)
$\mid P \,;\, Q$	(sequential composition)	$\mid PP \,;\, QQ$	
$\mid P \,\Box\, Q$	(choice)	$\mid PP \,\Box\, QQ$	
$\mid P \parallel Q$	(parallel composition)	$\mid PP \parallel QQ$	
$\mid SKIP$	(normal termination)	$\mid SKIPP$	
$\mid THROW$	(throw an interrupt)	$\mid THROWW$	
$\mid YIELD$	(yield to an interrupt)	$\mid YIELDD$	
$\mid P \rhd Q$	(interrupt handler)		
$\mid [PP]$	(transaction block)		

Fig. 1. cCSP syntax

points are inserted in a process through *YIELD*. For example, $(P \,;\, YIELD \,;\, Q)$ is willing to yield to an interrupt in between the execution of P, and Q. The basic way of constructing a compensable process is through a compensation pair $(P \div Q)$, which is constructed from two standard processes, where P is called the *forward* behaviour that executes during normal execution, and Q is called the associated compensation that is designed to compensate the effect of P when needed. The sequential composition of compensable processes is defined in such a way that the compensations of the completed tasks will be accumulated in reverse to the order of their original composition, whereas compensations from the compensable parallel processes will be placed in parallel. In this paper, we define only the asynchronous composition of processes, where processes interleave with each other during normal execution, and synchronise during termination. By enclosing a compensable process PP inside a transaction block $[\,PP\,]$, we get a complete transaction and the transaction block itself is a standard process. Successful completion of PP represents successful completion of the block. But, when the forward behaviour of PP throws an interrupt, the compensations are executed inside the block, and the interrupt is not observable from outside of the block. $SKIPP, THROWW$, and $YIELDD$ are the compensable counterpart of the corresponding standard processes and they are defined as follows:

$$\begin{aligned} SKIPP &= SKIP \div SKIP, \\ YIELDD &= YIELD \div SKIP \\ THROWW &= THROW \div SKIP \end{aligned}$$

To illustrate the use of cCSP, we present an example of a transaction for processing customer orders in a warehouse in Fig.2. The first step in the transaction is a compensation pair. The primary action of this pair is to accept the order and deduct the order quantity from the inventory database. The compensation action simply adds the order quantity back to the total in the inventory database. After an order is received from a customer, the order is packed for shipment, and a courier is booked to deliver the goods to the customer. The **PackOrder** process packs each of the items in the order in parallel. Each $PackItem$ activity can be compensated by a corresponding $UnpackItem$. Simultaneously with the packing of the order, a credit check is performed on the customer. The credit check is performed in parallel because it normally succeeds, and in this normal case the company does not wish to delay the order unnecessarily. In the case that a credit check fails, an interrupt is thrown causing the transaction to stop its execution, with the courier possibly having been booked and possibly some of the items having being packed. In case of failure, the semantics of the transaction block will ensure that the appropriate compensation activities will be invoked for those activities that already did take place.

$$\begin{aligned} \text{OrderTransaction} &= [\,\mathbf{ProcessOrder}\,] \\ \mathbf{ProcessOrder} &= (AcceptOrder \div RestockOrder)\,;\, \mathbf{FulfillOrder} \\ \mathbf{FulfillOrder} &= BookCourier \div CancelCourier \parallel \\ &\quad\ \mathbf{PackOrder} \parallel \\ &\quad\ CreditCheck\,;\ (\,Ok\,;\,SKIPP \\ &\qquad\qquad \Box\, NotOk\,;\,THROWW\,) \\ \mathbf{PackOrder} &= \|i \in Items \bullet (PackItem(i) \div UnpackItem(i)) \end{aligned}$$

Fig. 2. Warehouse order processing

III. Semantic Models

This section briefly outlines the trace and the operational semantics of cCSP.

A. Trace Semantics

A trace of a process records the history of behaviour up to some point. We show the operators on traces which are then lifted to operators on set of traces. Traces considered for cCSP are non-empty sets.

The trace of a standard process is of the form $s\langle\omega\rangle$ where $s \in \Sigma^*$ (Σ is alphabet of normal events) and $\omega \in \Omega$ ($\Omega = \{\checkmark, !, ?\}$), which means all traces end with any of the events in Ω, which is called a terminal event. The terminal events represent the termination of a process. Successful termination is shown by a $\checkmark$. Termination by either throwing or yielding an interrupt is shown by $!$ or $?$ respectively. In sequential composition $(p \,;\, q)$, the concatenated observable traces p and q, only when p terminates successfully,(ends with $\checkmark$), otherwise the trace is only p. The traces of two parallel processes are $p\langle\omega\rangle \| q\langle\omega'\rangle$ which corresponds to the set $(p \,|||\, q)$, the possible interleaving of traces of both processes and followed by $\omega\&\omega'$, the synchronisation of ω and ω'. The trace semantics of standard processes are shown in Fig. 3.

Compensable processes are comprised of *forward* and *compensation* behaviour. The traces of compensable processes are of pair of traces of the form $(s\langle\omega\rangle, s'\langle\omega'\rangle)$, where $s\langle\omega\rangle$ is the forward behaviour and $s'\langle\omega'\rangle$ is the compensation behaviour. In sequential composition, the forward traces correspond to the original forward behaviour and followed by the traces of the compensation. Traces of parallel composition are defined as the interleaving of forward traced and then follows the interleaving of compensation. The traces of a compensation pair are the traces of both of the processes of the pair when the forward process (P) terminate with a $\langle\checkmark\rangle$, otherwise the traces of the pair are the traces of the forward process followed by only a $\langle\checkmark\rangle$. The traces of a transaction block are only the traces of compensable processes inside the block when the process terminates with a $\langle\checkmark\rangle$, otherwise when the forward process inside the block terminates with a $\langle!\rangle$ the traces of

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

Atomic Action: For $A \in \Sigma \quad T(A) = \{\langle A, \checkmark\rangle\}$

Basic Processes:

$T(SKIP) = \{\langle\checkmark\rangle\},\ T(THROW) = \{\langle!\rangle\},\ T(YIELD) = \{\langle?\rangle, \langle\checkmark\rangle\}$

Sequential Composition:

$p\langle\checkmark\rangle\ ;\ q = p.q, \quad \text{and} \quad p\langle\omega\rangle\ ;\ q = p\langle\omega\rangle,\ \text{where}\ \omega \neq \checkmark$

$T(P\ ;\ Q) = \{p\ ;\ q \mid p \in P \wedge q \in Q\}$

Parallel Composition:

$p\langle\omega\rangle \| q\langle\omega'\rangle = \{r\langle\omega\&\omega'\rangle \mid r \in (p \mid\mid\mid q)\}$ where

ω	$!$	$!$	$!$	$?$	$?$	$\checkmark$
ω'	$!$	$?$	$\checkmark$	$?$	$\checkmark$	$\checkmark$
$\omega\&\omega'$	$!$	$!$	$!$	$?$	$?$	$\checkmark$

$T(P \| Q) = \{r \mid r \in (p \| q) \wedge p \in P \wedge q \in Q\}$

Interrupt Handler:

$p\langle!\rangle \rhd q = p.q \quad \text{and} \quad p\langle\omega\rangle \rhd q = p\langle\omega\rangle$ **where** $\omega \neq !$

$T(P \rhd Q) = \{p \rhd q \mid p \in P \ \wedge\ q \in Q\}$

Choice: $T(P \,\Box\, Q) = T(P) \cup T(Q)$

Transaction Block:

$[p\langle!\rangle, p'] = p.p' \quad \text{and} \quad [p\langle\checkmark\rangle, p'] = p\langle\checkmark\rangle$

$T([PP]) = \{[p, p'] \mid (p, p') \in PP\}$

Fig. 3. Trace semantics of standard processes

the block are the traces of the forward process followed by the traces of the compensation. Fig. 4 outlines the traces of compensable processes.

Basic Processes:

$T(SKIPP) = T(SKIP \div SKIP) = \{(\langle?\rangle, \langle\checkmark\rangle), (\langle\checkmark\rangle, \langle\checkmark\rangle)\}$

$T(THROWW) = T(THROWW \div SKIP) = \{(\langle?\rangle, \langle\checkmark\rangle), (\langle!\rangle, \langle\checkmark\rangle)\}$

$T(YIELDD) = T(YIELD \div SKIP) = \{(\langle?\rangle, \langle\checkmark\rangle)\}$

Compensation Pair:

$p\langle\checkmark\rangle \div q = (p\langle\checkmark\rangle, q) \quad \text{and} \quad p\langle\omega\rangle \div q = (p\langle\omega\rangle, \langle\checkmark\rangle)$ where $\omega \neq \checkmark$

$T(P \div Q) = \{(\langle?\rangle, \langle\checkmark\rangle)\} \ \cup \{p \div q \mid p \in P \wedge q \in Q\}$

Sequential Composition:

$(p\langle\checkmark\rangle, p')\ ;\ (q, q') = (pq, q'\ ;\ p')$

$(p\langle\omega\rangle, p')\ ;\ (q, q') = (p\langle\omega\rangle, p')$ where $\omega \neq \checkmark$

$T(PP\ ;\ QQ) = \{pp\ ;\ qq \mid pp \in PP \wedge qq \in QQ\}$

Parallel Composition:

$(p, p') \| (q, q') = \{(r, r') \mid r \in (p \| q) \wedge r' \in (p' \| q')\}$

$T(PP \| QQ) = \{rr \mid rr \in (pp \| qq) \wedge pp \in PP \wedge qq \in QQ\}$

Choice: $T(PP \,\Box\, PQ) = T(PP) \cup T(QQ)$

Fig. 4. Trace semantics of compensable processes

The following healthiness conditions declare that processes consist of some terminating or interrupting behaviour which ensures that the traces of processes are non-empty:

- $p\langle\checkmark\rangle \in T(P)$ or $p\langle!\rangle \in T(P)$, for some p
- $(p\langle\checkmark\rangle, p') \in T(PP)$ or $(p\langle!\rangle, p') \in T(PP)$, for some p, p'

B. Operational Semantics

By using labelled transition systems [5], the operational semantics specifies the relation between states of a program. Two types of transitions are define to present the transition relation of process terms: normal and terminal. A normal transition is defined by a normal event ($a \in \Sigma$) and a terminal transition is defined by a terminal event ($\omega \in \Omega$) .

For a standard process, a normal transition makes the transition of a process term from one state to its another state (P to P'). The terminal transition, on the other hand terminates a standard process to a null process (0):

$$P \xrightarrow{a} P', \qquad P \xrightarrow{\omega} 0$$

In sequential composition ($P\ ;\ Q$), the process Q can start only when the process P terminates successfully (with $\checkmark$). If P terminates with $!$ or $?$ the process Q will not start. In parallel composition each process can evolve independently and processes synchronise only on terminal events. The transition rules for standard processes are outlined in Fig. 5.

Atomic Action: $A \xrightarrow{A} SKIP \quad (A \in \Sigma)$

Basic Processes:

$SKIP \xrightarrow{\checkmark} 0,\ THROW \xrightarrow{!} 0,\ YILED \xrightarrow{?} 0,\ YIELD \xrightarrow{\checkmark} 0$

Sequential Composition:

$$\frac{P \xrightarrow{a} P'}{(P\ ;\ Q) \xrightarrow{a} (P'\ ;\ Q)} \qquad \frac{P \xrightarrow{\checkmark} 0 \wedge Q \xrightarrow{\alpha} Q'}{(P\ ;\ Q) \xrightarrow{\alpha} Q'} \qquad \frac{P \xrightarrow{\omega} 0}{(P\ ;\ Q) \xrightarrow{\omega} 0} \quad (\omega \neq \checkmark)$$

Parallel Composition:

$$\frac{P \xrightarrow{a} P'}{P \| Q \xrightarrow{a} P' \| Q} \qquad \frac{Q \xrightarrow{a} Q'}{P \| Q \xrightarrow{a} P \| Q'} \qquad \frac{P \xrightarrow{\omega} 0 \wedge Q \xrightarrow{\omega'} 0}{P \| Q \xrightarrow{\omega\&\omega'} 0}$$

where

ω	$!$	$!$	$!$	$?$	$?$	$\checkmark$
ω'	$!$	$?$	$\checkmark$	$?$	$\checkmark$	$\checkmark$
$\omega\&\omega'$	$!$	$!$	$!$	$?$	$?$	$\checkmark$

Choice: $\dfrac{P \xrightarrow{\alpha} P'}{P \,\Box\, Q \xrightarrow{\alpha} P'} \qquad \dfrac{Q \xrightarrow{\alpha} Q'}{P \,\Box\, Q \xrightarrow{\alpha} Q'}$

Interrupt handler:

$$\frac{P \xrightarrow{a} P'}{P \rhd Q \xrightarrow{a} P' \rhd Q} \qquad \frac{P \xrightarrow{!} 0 \wedge Q \xrightarrow{\alpha} Q'}{P \rhd Q \xrightarrow{\alpha} Q'} \qquad \frac{P \xrightarrow{\omega} 0}{P \rhd Q \xrightarrow{\omega} 0} \quad (\omega \neq !)$$

Transaction Block:

$$\frac{PP \xrightarrow{a} PP'}{[PP] \xrightarrow{a} [PP']} \qquad \frac{PP \xrightarrow{\checkmark} P}{[PP] \xrightarrow{\checkmark} 0} \qquad \frac{PP \xrightarrow{!} P \wedge P \xrightarrow{\alpha} P'}{[PP] \xrightarrow{\alpha} P'}$$

Fig. 5. Operational semantics of standard processes

For compensable processes, the normal transitions are same as standard processes. However, the terminal events terminate the forward behaviour of compensable processes, additionally, the compensation are stored for future reference.

$$PP \xrightarrow{a} PP', \qquad PP \xrightarrow{\omega} P \quad (P \text{ is the compensation})$$

In sequential composition ($PP\ ;\ QQ$), when PP terminates, its compensation (P) is stored and QQ starts to execute. In this scenario, we get an auxiliary construct ($\langle QQ, P\rangle$) where the processes have no particular operational relation between them. After termination of the process QQ, its compensation (Q) is accumulated in front of P i.e., ($Q\ ;\ P$). In the parallel composition, the main difference with the standard processes is that after termination of the forward behaviour the compensations are accumulated in parallel. The transition rules of compensable processes are summarised in Fig. 6.

A non-terminal event changes the state of the process inside the block. Successful completion of the forward process inside the block means completion of the whole block, but throwing a interrupt by the compensable process inside the block results the compensation to run. In compensation pair, after successful completion of the forward behaviour the compensation will be stored for future use, however, unsuccessful termination, i.e, terminates by $!$ or $?$ results an empty compensation (Fig. 5).

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

Choice:

$$\frac{PP \xrightarrow{a} PP'}{PP \,\Box\, QQ \xrightarrow{a} PP'} \quad \frac{QQ \xrightarrow{a} QQ'}{PP \,\Box\, QQ \xrightarrow{a} QQ'} \quad \frac{PP \xrightarrow{\omega} P}{PP \,\Box\, QQ \xrightarrow{\omega} P} \quad \frac{QQ \xrightarrow{\omega} Q}{PP \,\Box\, QQ \xrightarrow{\omega} Q}$$

Sequential Composition:

$$\frac{PP \xrightarrow{a} PP'}{PP \,;\, QQ \xrightarrow{a} PP' \,;\, QQ} \quad \frac{PP \xrightarrow{\checkmark} P \wedge QQ \xrightarrow{\omega} Q}{PP \,;\, QQ \xrightarrow{\omega} Q \,;\, P} \quad \frac{PP \xrightarrow{\omega} P}{PP \,;\, QQ \xrightarrow{\omega} P} \; (\omega \neq \checkmark)$$

$$\frac{PP \xrightarrow{\checkmark} P \wedge QQ \xrightarrow{a} QQ'}{PP \,;\, QQ \xrightarrow{a} \langle QQ', P\rangle} \quad \frac{QQ \xrightarrow{a} QQ'}{\langle QQ, P\rangle \xrightarrow{a} \langle QQ', P\rangle} \quad \frac{QQ \xrightarrow{\omega} Q}{\langle QQ, P\rangle \xrightarrow{\omega} Q \,;\, P}$$

Parallel Composition:

$$\frac{PP \xrightarrow{a} PP'}{PP \parallel QQ \xrightarrow{a} PP' \parallel QQ} \quad \frac{QQ \xrightarrow{a} QQ'}{PP \parallel QQ \xrightarrow{a} PP \parallel QQ'} \quad \frac{PP \xrightarrow{\omega} P \wedge QQ \xrightarrow{\omega'} Q}{PP \parallel QQ \xrightarrow{\omega\&\omega'} P \parallel Q}$$

Compensation Pair:

$$\frac{P \xrightarrow{a} P'}{P \div Q \xrightarrow{a} P' \div Q} \quad \frac{P \xrightarrow{\checkmark} 0}{P \div Q \xrightarrow{\checkmark} Q} \quad \frac{P \xrightarrow{\omega} 0}{P \div Q \xrightarrow{\omega} SKIP} \; (\omega \neq \checkmark)$$

Fig. 6. Operational semantics of compensable processes

IV. Relating Semantic Models

In this section we describe the steps to derive a relationship between the two semantic models of cCSP. We follow a systematic approach to derive the relationship where traces are first extracted from the transition rules and prove that the extracted traces correspond to the original trace definition. The steps of deriving the semantic relation are shown in Fig. 7.

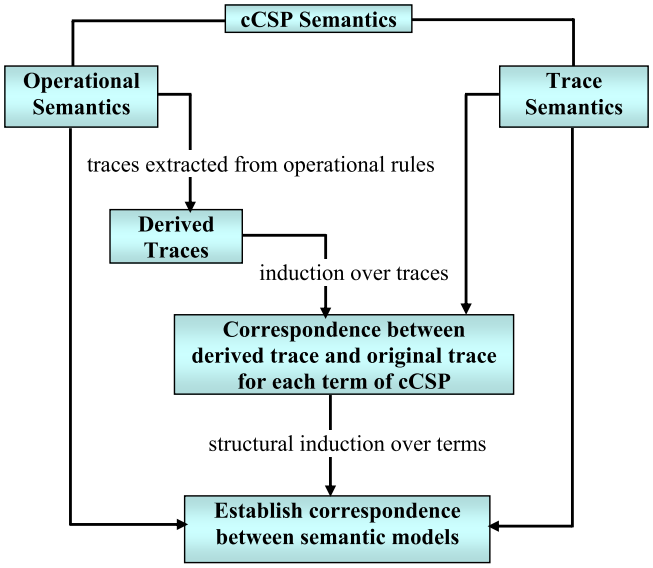


Fig. 7. Steps to derive relationship between semantic models

The operational semantics leads to lifted transition relations labelled by sequences of events. This is defined recursively. For a standard process P,

$$P \xrightarrow{\langle\omega\rangle} Q \quad = \quad P \xrightarrow{\omega} Q$$
$$P \xrightarrow{\langle a\rangle t} Q \quad = \quad \exists P' \cdot P \xrightarrow{a} P' \;\wedge\; P' \xrightarrow{t} Q$$

The derived traces of a standard process P is defined as $DT(P)$. Let $t \in DT(P)$, then we get the following definition,

$$t \in DT(P) \quad = \quad P \xrightarrow{t} 0 \tag{1}$$

Compensable processes have both forward and compensation behaviour. A compensable process is defined as a pair of traces. Hence, it is required to extract traces from both forward and compensation behaviour. The forward behaviour of a compensable process PP is defined as follows:

$$PP \xrightarrow{t} R \quad (t \text{ ends with } \omega)$$

where t is the trace of the forward behaviour. R is the attached compensation. The behaviour of compensation is similar to standard processes and by reusing that we get the following definition:

$$PP \xrightarrow{(t,t')} 0 \quad = \quad \exists R \cdot PP \xrightarrow{t} R \;\wedge\; R \xrightarrow{t'} 0$$

where t' is the trace of the compensation. For a compensable process PP, the derived traces $DT(PP)$ is defined as follows:

$$(t,t') \in DT(PP) \quad = \quad PP \xrightarrow{(t,t')} 0$$

By using the definition of derived traces and the original traces we state the following theorem to define the relationship between the semantic models,

Theorem 1: For any standard process term P, where $P \neq 0$

$$DT(P) \;=\; T(P)$$

For any compensable process terms PP, where $PP \neq 0$ and does not contain the term $\langle PP, P\rangle$,

$$DT(PP) \quad = \quad T(PP)$$

Traces are extracted for each term of the language, and its correspondence is shown with the corresponding traces in the trace semantics. Assume P and Q are standard process terms, then for all the operators, we prove that

$$t \in DT(P \otimes Q) \quad = \quad t \in T(P \otimes Q) \tag{2}$$

For each such operator $\otimes$, the proof is performed by induction over traces. In the proof we assume that, $DT(P) = T(P)$ and $DT(Q) = T(Q)$.

We follow similar style for compensable processes. Assuming $DT(PP) = T(PP)$ and $DT(QQ) = T(QQ)$ we show that,

$$(t,t') \in DT(PP \otimes QQ) \quad = \quad (t,t') \in T(PP \otimes QQ) \tag{3}$$

In the following sections we outline the proof steps showing the correspondence in (2) and (3) for both standard and compensable process terms.

A. Standard Processes

Sequential Composition: By using (2) the relationship between the semantic models is derived by showing that,

$$t \in DT(P \,;\, Q) \;=\; t \in T(P \,;\, Q)$$

From (1) we get the derived traces of the sequential composition,

$$t \in DT(P \,;\, Q) \quad = \quad (P \,;\, Q) \xrightarrow{t} 0$$

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

We also expand the definition of trace semantics as follows:

$$\begin{aligned} & t \in T(P\ ;Q) \\ & = \exists p,q \cdot t = (p\ ;\ q)\ \wedge\ p \in T(P)\ \wedge\ q \in T(Q) \\ & = \exists p,q \cdot t = (p\ ;\ q)\ \wedge\ p \in DT(P)\ \wedge\ q \in DT(Q) \\ & = \exists p,q \cdot t = (p\ ;\ q)\ \wedge\ P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \end{aligned}$$

Finally, from the above definitions of traces, the following lemma is formulated for the sequential composition of standard processes:

Lemma 1:

$$(P\ ;\ Q) \xrightarrow{t} 0\ =\ \exists p,q \cdot t = (p\ ;\ q)\ \wedge\ P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0$$

The lemma is proved by applying induction over the trace t, where $t = \langle\omega\rangle$ is considered as the base case, and $t = \langle a\rangle t$ is considered as the inductive case. To support the proof of the lemma, two equations are derived from the transition rules. These derived equations are based on the event by which the transition rules are defined:

$$\begin{aligned} (P\ ;\ Q) \xrightarrow{\omega} 0 \quad &= \quad P \xrightarrow{\checkmark} 0\ \wedge\ Q \xrightarrow{\omega} 0 \\ &\vee \quad P \xrightarrow{\omega} 0\ \wedge\ \omega \neq \checkmark \\ (P\ ;\ Q) \xrightarrow{a} R \quad &= \quad \exists P' \cdot P \xrightarrow{a} P'\ \wedge\ R = (P'\ ;\ Q) \\ &\vee \quad P \xrightarrow{\checkmark} 0\ \wedge\ Q \xrightarrow{a} R \end{aligned}$$

Proof:

Basic step: $t = \langle\omega\rangle$

$(P\ ;\ Q) \xrightarrow{\langle\omega\rangle} 0 \quad = \quad (P\ ;\ Q) \xrightarrow{\omega} 0$

"From transition rules sequential composition"

$$= \quad P \xrightarrow{\checkmark} 0\ \wedge\ Q \xrightarrow{\omega} 0 \qquad (4)$$

$$\vee \quad P \xrightarrow{\omega} 0\ \wedge\ \omega \neq \checkmark \qquad (5)$$

From (4)

$$\begin{aligned} & P \xrightarrow{\checkmark} 0\ \wedge\ Q \xrightarrow{\omega} 0 \\ = \ & \exists p,q \cdot p = \langle\checkmark\rangle\ \wedge\ q = \langle\omega\rangle\ \wedge\ \langle\omega\rangle = (p\ ;\ q) \\ \wedge \ & P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \\ = \ & \exists p,q \cdot \langle\omega\rangle = (p\ ;\ q)\ \wedge\ p = \langle\checkmark\rangle \\ \wedge \ & P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \end{aligned}$$

From (5)

$$\begin{aligned} & P \xrightarrow{\omega} 0\ \wedge\ \omega \neq \checkmark \\ = \ & \exists p,q \cdot p = \langle\omega\rangle\ \wedge\ \omega \neq \checkmark\ \wedge\ \langle\omega\rangle = (p\ ;\ q) \\ \wedge \ & P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \\ = \ & \exists p,q \cdot \langle\omega\rangle = (p\ ;\ q)\ \wedge\ p \neq \langle\checkmark\rangle \\ \wedge \ & P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \end{aligned}$$

Therefore, for $t = \langle\omega\rangle$ from (4) and (5)

$$\begin{aligned} & \exists p,q \cdot \langle\omega\rangle = (p\ ;\ q)\ \wedge\ p = \langle\checkmark\rangle \\ & \wedge P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \\ \vee \ & \exists p,q \cdot \langle\omega\rangle = (p\ ;\ q)\ \wedge\ p \neq \langle\checkmark\rangle \\ & \wedge P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \\ = \ & \exists p,q \cdot \langle\omega\rangle = (p\ ;\ q)\ \wedge\ P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \end{aligned}$$

Inductive step: $t = \langle a\rangle t$

$P\ ;Q \xrightarrow{\langle a\rangle t} 0 \quad = \quad \exists R \cdot (P\ ;\ Q) \xrightarrow{a} R\ \wedge\ R \xrightarrow{t} 0$

"From operational rules"

$$= \quad \exists P' \cdot P \xrightarrow{a} P'\ \wedge\ (P'\ ;\ Q) \xrightarrow{t} 0 \qquad (6)$$

$$\vee\ \exists Q' \cdot P \xrightarrow{\checkmark} 0\ \wedge\ Q \xrightarrow{a} Q'\ \wedge\ Q' \xrightarrow{t} 0 \qquad (7)$$

From (6)

$$\begin{aligned} & \exists P' \cdot P \xrightarrow{a} P'\ \wedge\ (P'\ ;\ Q) \xrightarrow{t} 0 \\ = \ & \text{"Inductive hypothesis"} \\ & \exists P' \cdot P \xrightarrow{a} P'\ \wedge\ \exists p',q \cdot t = (p'\ ;\ q) \\ \wedge \ & P' \xrightarrow{p'} 0\ \wedge\ Q \xrightarrow{q} 0 \\ = \ & \text{"Combining existential quantifications"} \\ & \exists p',q \cdot t = (p'\ ;\ q)\ \wedge\ P \xrightarrow{\langle a\rangle p'} 0\ \wedge\ Q \xrightarrow{q} 0 \\ = \ & \text{"Using trace rule } \langle a\rangle t = \langle a\rangle(p'\ ;\ q) = (\langle a\rangle p')\ ;\ q\text{"} \\ & \exists p',q \cdot \langle a\rangle t = (\langle a\rangle p'\ ;\ q)\ \wedge\ P \xrightarrow{\langle a\rangle p'} 0\ \wedge\ Q \xrightarrow{q} 0 \\ = \ & \exists p,q \cdot p = \langle a\rangle p'\ \wedge\ \langle a\rangle t = (p\ ;\ q) \\ \wedge \ & P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \end{aligned}$$

From (7)

$$\begin{aligned} & \exists Q' \cdot P \xrightarrow{\checkmark} 0\ \wedge\ Q \xrightarrow{a} Q'\ \wedge\ Q' \xrightarrow{t} 0 \\ = \ & P \xrightarrow{\checkmark} 0\ \wedge\ Q \xrightarrow{\langle a\rangle t} 0 \\ = \ & \exists p,q \cdot p = \langle\checkmark\rangle\ \wedge\ q = \langle a\rangle t\ \wedge\ \langle a\rangle t = (p\ ;\ q) \\ \wedge \ & P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \\ = \ & \exists p,q \cdot \langle a\rangle t = (p\ ;\ q)\ \wedge\ p = \langle\checkmark\rangle \\ \wedge \ & P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \end{aligned}$$

Therefore for $t = \langle a\rangle t$, from (6) $\vee$ (7)

$$\begin{aligned} & \exists p,q \cdot p = \langle a\rangle p'\ \wedge\ \langle a\rangle t = (p\ ;\ q)\ \wedge\ P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \\ \vee \ & \exists p,q \cdot p = \langle\checkmark\rangle\ \wedge\ \langle a\rangle t = (p\ ;\ q)\ \wedge\ P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \\ = \ & \text{"Combining existential quantifications"} \\ & \exists p,q \cdot (p = \langle\checkmark\rangle\ \vee\ p = \langle a\rangle p')\ \wedge\ \langle a\rangle t = (p\ ;\ q) \\ & \wedge\ P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \\ = \ & \exists p,q \cdot \langle a\rangle t = (p\ ;\ q)\ \wedge\ P \xrightarrow{p} 0\ \wedge\ Q \xrightarrow{q} 0 \end{aligned}$$

■

This completes the proof of the lemma. We follow the same approach to prove other lemmas in the rest of the paper.

Parallel Composition: The parallel composition of two processes is defined to be the interleaving of their observable events followed by the synchronisation of their terminal events. For example, considering asynchronous actions, $A \parallel B$ can execute A followed by B or B followed by A. For traces p and q we write $p \mid\mid\mid q$ to denote the set of interleaving of p and q and it follows the following definition:

$$\begin{aligned} \langle\rangle \in p \mid\mid\mid q \quad &= \quad p = \langle\rangle\ \wedge\ q = \langle\rangle \\ \langle a\rangle t \in p \mid\mid\mid q \quad &= \quad \exists p' \cdot p = \langle a\rangle p'\ \wedge\ t \in p' \mid\mid\mid q \\ &\vee \quad \exists q' \cdot q = \langle a\rangle q'\ \wedge\ t \in p \mid\mid\mid q' \end{aligned}$$

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

By following similar steps as sequential composition, we define the following lemma for parallel composition:

Lemma 2:

$$(P \parallel Q) \xrightarrow{t} 0 \;=\; \exists p,q \cdot t \in (p \parallel q) \wedge P \xrightarrow{p} 0 \wedge Q \xrightarrow{q} 0$$

We derive two supporting equation from the transition rules of parallel composition:

$$\begin{aligned} P \parallel Q \xrightarrow{a} R &= P \xrightarrow{a} P' \wedge R = P' \parallel Q \\ &\vee \quad Q \xrightarrow{a} Q' \wedge R = P \parallel Q' \\ P \parallel Q \xrightarrow{\omega} 0 &= P \xrightarrow{\omega 1} 0 \wedge Q \xrightarrow{\omega 2} 0 \wedge \omega \in \omega 1 \& \omega 2 \end{aligned}$$

Proof: The proof of the base case is trivial and omitted from the presentation. The inductive case is described here:

$$\begin{aligned} & (P \parallel Q) \xrightarrow{\langle a \rangle t} 0 \\ = \;& \exists R \cdot (P \parallel Q) \xrightarrow{\langle a \rangle} R \wedge R \xrightarrow{t} 0 \\ = \;& \text{“Using the operational rules”} \\ & \exists P' \cdot P \xrightarrow{a} P' \wedge (P' \parallel Q) \xrightarrow{t} 0 \\ \vee \;& \exists Q' \cdot Q \xrightarrow{a} Q' \wedge (P \parallel Q') \xrightarrow{t} 0 \\ = \;& \text{“Inductive hypothesis”} \\ & \exists P' \cdot P \xrightarrow{a} P' \wedge \exists p', q \cdot t \in (p' \parallel q) \\ \wedge \;& P' \xrightarrow{p'} 0 \wedge Q \xrightarrow{q} 0 \\ \vee \;& \exists Q' \cdot Q \xrightarrow{a} Q' \wedge \exists p, q' \cdot t \in (p \parallel q') \\ \wedge \;& P \xrightarrow{p} 0 \wedge Q' \xrightarrow{q'} 0 \\ = \;& \text{“Combining existential quantifications”} \\ = \;& \exists p', q \cdot t \in (p' \parallel q) \wedge P \xrightarrow{\langle a \rangle p'} 0 \wedge Q \xrightarrow{q} 0 \\ \vee \;& \exists p, q' \cdot t \in (p \parallel q') \wedge P \xrightarrow{p} 0 \wedge Q \xrightarrow{\langle a \rangle q'} 0 \\ = \;& \exists p, q \cdot p = \langle a \rangle p' \wedge t \in (p' \parallel q) \\ & \wedge P \xrightarrow{p} 0 \wedge Q \xrightarrow{q} 0 \\ \vee \;& \exists p, q \cdot q = \langle a \rangle q' \wedge t \in (p \parallel q') \\ & \wedge P \xrightarrow{p} 0 \wedge Q \xrightarrow{q} 0 \\ = \;& \text{“Combining”} \\ & \exists p, q \cdot (p = \langle a \rangle p' \wedge t \in (p' \parallel q) \vee q = \langle a \rangle q' \\ & \wedge t \in (p \parallel q')) \wedge P \xrightarrow{p} 0 \wedge Q \xrightarrow{q} 0 \\ = \;& \text{“By the definition the interleaving of traces”} \\ & \exists p, q \cdot \langle a \rangle t \in (p \parallel q) \wedge P \xrightarrow{p} 0 \wedge Q \xrightarrow{q} 0 \end{aligned}$$

■

B. Compensable Processes

Sequential Composition: For compensable processes *PP* and *QQ*, let $(t,t') \in DT(PP \;;\; QQ)$ and according to trace derivation rule we get

$$(t,t') \in DT(PP \;;\; QQ) = \exists R \cdot (PP \;;\; QQ) \xrightarrow{t} R \wedge R \xrightarrow{t'} 0$$

The following lemma is stated to define the relationship for the lifted forward behaviour of sequential composition of compensable processes:

Lemma 3:

$$\begin{aligned} (PP \;;\; QQ) \xrightarrow{t} R &= \exists P,Q,p,q \cdot t = (p \;;\; q) \\ &\wedge \quad PP \xrightarrow{p} P \wedge QQ \xrightarrow{q} Q \\ &\wedge \quad R = COND(last(p) = \checkmark, (Q \;;\; P), P) \\ \textit{Where}, \quad & COND(true, e1, e2) = e1 \\ & COND(false, e1, e2) = e2 \end{aligned}$$

COND expression is used to state that when process *PP* terminates successfully (terminate by ✓), compensation from both *PP* and *QQ* are accumulated in reverse order, otherwise only compensation from *PP* is stored. The following equations are derived from the transition rules to support the proof of the above lemma.

$$\begin{aligned} (PP \;;\; QQ) \xrightarrow{a} RR &= PP \xrightarrow{a} PP' \wedge RR = (PP' \;;\; QQ) \\ &\vee \quad PP \xrightarrow{\checkmark} P \wedge QQ \xrightarrow{a} QQ' \\ &\wedge \quad R = \langle QQ', P \rangle \\ (PP \;;\; QQ) \xrightarrow{\omega} R &= PP \xrightarrow{\checkmark} P \wedge QQ \xrightarrow{a} Q \wedge R = (Q \;;P) \\ &\vee \quad PP \xrightarrow{\omega} P \wedge \omega \neq \checkmark \wedge R = P \end{aligned}$$

In the inductive case of the lemma we get the following intermediate step involving the auxiliary construct $\langle QQ, P \rangle$.

$$\begin{aligned} PP \;;\; QQ \xrightarrow{\langle a \rangle t} R &= \exists RR \cdot PP \;;QQ \xrightarrow{a} RR \wedge RR \xrightarrow{t} R \\ &= \exists PP' \cdot PP \xrightarrow{a} PP' \wedge PP' \;;QQ \xrightarrow{t} R \\ &\vee \quad \exists P, QQ' \cdot PP \xrightarrow{\checkmark} P \wedge QQ \xrightarrow{a} QQ' \\ &\wedge \quad \langle QQ', P \rangle \xrightarrow{t} R \end{aligned} \tag{8}$$

To deal with this we need another lemma which will support the removal of auxiliary construct in (8). This lemma considers the situation where the forward behaviour of the first process of sequential composition is terminated with ✓ and its compensation is stored and the second process of the composition has started. Here to mention that *t* in (8) above is a complete trace.

Lemma 4:

$$\langle QQ, P \rangle \xrightarrow{t} R = \exists Q \cdot QQ \xrightarrow{t} Q \wedge R = (Q \;;P)$$

The lemma is proved by induction over traces. By using this lemma, we prove Lemma 3 by following the similar approach of applying induction over traces.

Parallel Composition: Let $(t,t') \in DT(PP \parallel QQ)$ By using the trace derivation rule we get,

$$(t,t') \in DT(PP \parallel QQ) = \exists R \cdot (PP \parallel QQ) \xrightarrow{t} R \wedge R \xrightarrow{t'} 0$$

We then define the following lemma to establish the semantic correspondence for parallel composition of compensable processes:

Lemma 5:

$$\begin{aligned} (PP \parallel QQ) \xrightarrow{t} R &= \exists P,Q,p,q \cdot t \in (p \parallel q) \\ &\wedge PP \xrightarrow{p} P \wedge QQ \xrightarrow{q} P \wedge R = P \parallel Q \end{aligned}$$

The lemma is proved by using induction over traces similar to other lemmas.

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

Compensation Pair: A compensation pair $(P \div Q)$ consists of two standard processes: a standard process (P) and its compensation (Q). The semantics of compensation pair is defined in such a way that the behaviour of the compensation Q is augmented only with successfully completed forward behaviour of P, otherwise, the compensation is empty. For a compensation pair, we show that

$$(t,t') \in DT(P \div Q) \quad = \quad (t,t') \in T(P \div Q)$$

To prove the semantic correspondence between the semantics model, we state the following lemma:

Lemma 6:

$$(P \div Q) \xrightarrow{(t,t')} 0 \;=\; \exists p,q \cdot (t,t') = (p \div Q) \wedge P \xrightarrow{p} 0 \wedge Q \xrightarrow{q} 0$$

The lemma is proved by induction as previous lemmas. To support the inductive proof, the following two equations are derived from the transitions rules shown earlier,

$$\begin{aligned}(P \div Q) \xrightarrow{\omega} R \quad &= \quad P \xrightarrow{\checkmark} 0 \;\wedge\; R = Q \\ &\vee \quad P \xrightarrow{\omega} 0 \;\wedge\; \omega \neq \checkmark \;\wedge\; R = SKIP \\ (P \div Q) \xrightarrow{a} RR \quad &= \quad P \xrightarrow{a} P' \;\wedge\; RR = P' \div Q\end{aligned}$$

Unlike the lemmas defined earlier for compensable processes, Lemma 6 includes the traces of both forward and compensation behaviour. The following trace rules for the compensation pair are used in the proof of the lamma:

$$\begin{aligned}\text{when } p = p'\langle\checkmark\rangle (t,t') \quad &= \quad (p'\langle\checkmark\rangle \div q) \;=\; (p,q) \\ \text{when } p = p'\langle\omega\rangle \wedge \omega \neq \checkmark (t,t') \quad &= \quad (p'\langle\omega\rangle \div q) \;=\; (p,\langle\checkmark\rangle)\end{aligned}$$

Transaction Block: Transaction block is a standard process. We let $t \in DT([PP])$ and by following the trace derivation rule we get

$$t \in DT([PP]) \quad = \quad [PP] \xrightarrow{t} 0$$

The semantic correspondence is then derived by proving the following lemma:

Lemma 7:

$[PP] \xrightarrow{t} 0 = \exists p,p' \cdot t = [p,p'] \;\wedge\; PP \xrightarrow{p,p'} 0$

The operational semantics provide us the following equations to support the proof of the above lemma.

$$\begin{aligned}[PP] \xrightarrow{a} R \quad &= \quad PP \xrightarrow{a} PP' \;\wedge\; R = [PP'] \\ &\vee \quad PP \xrightarrow{!} P \;\wedge\; P \xrightarrow{a} P' \;\wedge\; R = P' \\ [PP] \xrightarrow{\omega} 0 \quad &= \quad PP \xrightarrow{\checkmark} P \;\wedge\; P \xrightarrow{p'} 0 \\ &\vee \quad PP \xrightarrow{!} P \;\wedge\; P \xrightarrow{\omega} 0\end{aligned}$$

The block operator runs the compensation of a terminating forward behaviour and discards the compensation of successfully completed forward behaviour. It removes the traces of an yielding forward behaviour.

We left two operators from the correspondence proof presented here. First one is the choice operator ($P \,\Box\, Q$). The trace of choice is the union of their traces and the operational rules shows that either process (P or Q) can evolve independently. Correspondence proof of this operator is trivial. Another operator that was left is interrupt handler ($P \rhd Q$). It is quite similar to standard sequential composition except that the flow of control from the first to the second process is caused by a throw (!) rather than a ✓ and hence, showing its correspondence proof would be repetitive.

V. Lessons Learned

We have adopted a systematic approach to show the correspondence between the two semantic models of cCSP. Traces are derived from the operational rules and then applying induction over the traces we showed the correspondence. Due to the way of defining operational rules the trace derivation was done easily. We used labelled transition system to define the operational rules. In [6] operational rules are defined for a similar language as ours but same symbol is used to define the labels of different transition rules. However, we used special symbols for different kinds of transitions. Transition between states are caused by two kinds of events: normal and terminal and we used these events as labels in our transition rules. The advantage of this approach of defining labels is that these labels are the traces of the transition and we can then derive these traces from the transition rules.

The trace operators play a significant role in defining the lemmas as well as in the correspondence proofs. The operators are used both at the trace levels and at the process levels. All the lemmas defined in this chapter have a common pattern applicable to both standard and compensable processes. For example, for standard processes P and Q, and their traces p and q, the lemmas for all the operators are defined as follows:

$$(P \otimes Q) \xrightarrow{t} 0 \;=\; \exists p,q \cdot t = (p \otimes q) \wedge P \xrightarrow{p} 0 \wedge Q \xrightarrow{q} 0$$

(for parallel operator use $t \in (p \otimes q)$ instead of $t = (p \otimes q)$)

Similar definitions are also given for the forward behaviour of compensable processes. The use of operators at both trace and process levels allow us to apply appropriate rules for the operators (rules for terminal and observable events from operational and trace semantics).

The correspondence was proved by using structural induction. First, the induction was applied on process terms of the language and then on the derived traces. The lower level induction which is on traces support the induction on upper level which is on process terms

VI. Related Work

The semantic correspondence presented here is based on the technique of applying structural induction. A similar approach is also applied by S. Schneider [7], where an equivalence relation was established between the operational and denotational semantics of timed CSP [8][9]. Operational rules are defined for timed CSP and then timed traces and refusals are extracted from the transition rules of a program, and it is shown that the pertinent information corresponds to the semantics obtained from the denotational semantic function. By applying structural induction over the terms of timed CSP, it was proved that the behaviour of the transition system is identical to those provided by the denotational semantics.

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

A similar problem was also investigated in [10], where a metric structure was employed to relate the operational and denotational models of a given language. In order to relate the semantic models it was proved that the two models coincide. The denotational models were extended and structural induction was applied over the terms of the language to relate the semantic models.

Other than using induction, Hoare and He [11] presented the idea of unifying different programming paradigms and showed how to derive operational semantics from its denotational presentation of a sequential language. They derive algebraic laws from the denotational definition and then derive the operational semantics from the algebraic laws. Similar to our work, Huibiao *et al.* [12] derived denotational semantics from operational semantics for a subset of Verilog [13]. However the derivation was done in a different way than our method where the authors defined transitional condition and phase semantics from the operational semantics. The denotational semantics are derived from the sequential composition of the phase semantics. The authors also derived operational semantics from denotational semantics [14].

Unlike our approach, the unification between the two semantics was shown in [15] by extending the operational semantics to incorporate the denotational properties. The equivalence was shown for a language having simple models without any support for concurrency. Similar problem was also investigated in [16] for a simple sequential language, which support recursion and synchronisation in the form of interleaving. The relation between operational and denotational semantics is obtained via an intermediate semantics.

VII. Concluding Remarks

It is of great importance to have the description of both operational and denotational semantics. Having both of the semantics we need to establish a relationship between these two. Demonstrating the relationship between these two semantics of the same language ensures the consistency of the whole semantic description of the language.

The main contribution of this paper is to show the correspondence between the operational semantics and the trace semantics of a subset of cCSP language. The correspondence is shown by deriving the traces from the operational rules and then applying the induction over the derived traces. Two level of induction is applied. In one level induction is applied over the operational rules and in the next level induction is applied over the derived traces.

The correspondence shown here are completely done by hand which is error prone and there are strong possibilities to miss some of the important parts during the proof. As part of the future work our goal is to use an automated/mechanized prover which will help us to use the similar approach that we followed here i.e, mathematical induction, and at the same time prove the theorems automatically. Among several tools we are currently using PVS (Prototype Verification System) [17] for our purpose. The specification language of PVS is based on classical, typed, high order logic and contains the constructs intended to ease the natural development of specification. The PVS proof checker is interactive and provides powerful basic commands and a mechanism for building re-usable strategies based on these.

The parallel operator of cCSP does not support synchronization on normal events. Synchronization of events is significant for the development of a language. Currently we are working on adding synchronization to cCSP. Adding synchronization and then using mechanized theorem prover for showing the correspondence will strengthen the formal foundation of the language.

References

[1] M. Butler, T. Hoare, and C. Ferreira, “A trace semactics for long-running transaction,” in *Proceedings of 25 Years of CSP*, ser. LNCS, A. Abdallah, C. Jones, and J. Sanders, Eds., vol. 3525. London: Springer-Verlag, 2004.

[2] C. Hoare, *Communicating Sequential Process*. Prentice Hall, 1985.

[3] J. Gray and A. Reuter, *Transaction Processing : Concepts and Techniques*. Morgan Kaufmann Publishers, 1993.

[4] M. Butler and S. Ripon, “Executable semantics for compensating CSP,” in *WS-FM 2005*, ser. LNCS, M. Bravetti, L. Kloul, and G. Zavattaro, Eds., vol. 3670. Versailles, France: Springer-Verlag, September 1-3 2005, pp. 243–256.

[5] G. D. Plotkin, “A structural approach to operational semantics.” Aarhus University, Computer Science Department, Tech. Rep. DAIMI FN-19, September 1981.

[6] R. Bruni, H. Melgratti, and U. Montanari, “Theoretical foundations for compensations in flow composition languages,” in *POPL*, 12-14 January 2005, pp. 209–220.

[7] S. Schneider, “An operational semantics for timed CSP,” *Journal of Information and computing*, vol. 116, no. 2, pp. 193–213, 1995.

[8] G. M. Reed and A. W. Roscoe, “A timed model for communicating sequential processes,” *Theoretical Computer Science*, vol. 58, no. 1-3, pp. 249–261, June 1988.

[9] S. Schneider, J. Davies, D. M. Jackson, G. M. Reed, J. N. Reed, and A. W. Roscoe, “Timed CSP: Theory and practice,” in *REX Workshop*, ser. LNCS, vol. 600, 1991, pp. 640–675.

[10] F. van Breugel, “An introduction to metric semantics: operational and denotational models for programming and specification languages,” *Theoretical Computer Science*, vol. 258, no. 1-2, pp. 1–98, May 2001.

[11] C. Hoare and H. Jifeng, *Unifying Theories of Programming*. Prentice Hall International Series in Computer Science, 1998.

[12] H. Zhu, J. P. Bowen, and J. He, “From operational semantics to denotational semantics for Verilog,” in *CHARME 2001*, ser. LNCS, T. Margaria and T. F. Melham, Eds., vol. 2144, 2001, pp. 449–466.

[13] M. Gordon, “The semantic challenge of Verilog HDL,” in *Proceedings of the 10th Annual IEEE Symposium on Logic in Computer Science (LICS ’95:)*. IEEE Computer Society, June 1995, pp. 136–145.

[14] H. Zhu, J. P. Bowen, and J. He, “Deriving operational semantics from denotational semantics for Verilog,” in *8th Asia-Pacific Software Engineering Conference (APSEC 2001)*. IEEE Computer Society, 4-7 Dec 2001, pp. 177 – 184.

[15] S. F. Smith, “From operational to denotational semantics,” in *Proceedings of the 7th International Conference on Mathematical Foundations of Programming Semantics*, ser. LNCS, vol. 598, 1992, pp. 54–76.

[16] J.-J. C. Meyer and E. Vink, *On Relating Denotational and Operational Semantics for Programming Languages with Recursion and Concurrency*, ser. Open Problems in Topology. Elsevier, 1990, ch. 24, pp. 387–406.

[17] S. Owre, J. Rushby, and N. Shankar, “PVS: A Prototype Verification System,” in *11th International Conference on Automated Deduction (CADE)*, ser. Lecture Notes in Artificial Intelligence, D. Kapur, Ed., vol. 607. Springer-Verlag, June 1992, pp. 748–752.

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

The Importance Analysis of Use Case Map with Markov Chains

Yaping Feng
School of Computer engineering, Kumoh National Institute of Technology, Korea

Lee-Sub Lee
School of Computer engineering, Kumoh National Institute of Technology, Korea

***Abstract*—UCMs (Use Case Maps) model describes functional requirements and high-level designs with causal paths superimposed on a structure of components. It could provide useful resources for software acceptance testing. However until now statistical testing technologies for large scale software is not considered yet in UCMs model. Thus if one applies UCMs model to a large scale software using traditional coverage-based exhaustive tasting, then it requires too much costs for the quality assurance. Therefore this paper proposes an importance analysis of UCMs model with Markov chains. With this approach not only highly frequently used usage scenarios but also important objects such as components, responsibilities, stubs and plug-ins can also be identified from UCMs specifications. Therefore careful analysis, design, implementation and efficient testing could be possible with the importance of scenarios and objects during the full software life cycle. Consequently product reliability can be obtained with low costs. This paper includes an importance analysis method that identifies important scenarios and objects and a case study to illustrate the applicability of the proposed approach.**

Keywords- Use Case Maps; Markov chain; Usability Testing.

I. INTRODUCTION

UCMs (Use Case Maps) [1, 2, 3] is a set of semi-formal notations for describing scenarios of a system. This notation is being standardized as a part of the URN (User Requirement Notation) that is the most recent addition to ITU-T's (International Telecommunication Union-Telecommunication) family of languages. UCMs model provides a scenario-based model for a system. And it has been successfully applied to wide range of systems, including telecommunication systems [4], distributed systems [1] and etc. Furthermore, Daniel Amyot applied UCMs model for customer-oriented acceptance tests for Web applications [5]. But existing research works have not considered statistical testing method yet. Traditional coverage-based exhaustive testing might be impractical for the large scale software products because of the size as well as the generally uneven distribution of problems and usage frequencies in different areas and product components.

The general solution of testing for large software products uses product reliability goals as a stopping criterion. The use of the criterion requires testing to be performed under an environment that resembles actual usage by target customers so that realistic reliability assessment can be obtained, resulting in the so-called SUT (Statistical Usage Testing) [3].

A prerequisite to such statistical testing and reliability analysis strategies is collection of usage information and construction of corresponding usage models. UCMs describe functional requirements and high-level designs with causal scenarios superimposed on a structure of components. Furthermore individual scenarios from UCMs can be generated automatically from related tools [6]. Hence, it is straightforward to collect usage information from UCMs to construct usage model.

This paper proposes an importance analysis of UCMs model with Markov chains which applies statistical technique to UCMs model. If one simply applies traditional statistical techniques to usecase model, only highly frequently used usage scenarios can be identified and tested more thoroughly than less frequently used ones. But the approach of this paper is based on UCMs model which includes more objects like components, responsibilities, stubs and plug-ins. So this paper has proposed mechanisms and object model of UCMs which can calculate the importance of these objects. Thus efficient testing can be given to both highly frequently used usage scenarios and important objects to assure and maximize the product reliability from a customer's perspective with low cost. Since the combination of UCMs model and statistical technique can identify much more important information, it will be very beneficial for quality assurance engineering.

This paper is organized as follows. Related works are discussed in section 2. Section 3 describes the background concepts of UCM scenario models and Markov chain usage models. An importance analysis of UCM with Markov chains is presented in section 4 with a case study. Conclusions and future works are discussed in section 5.

II. RELATED WORKS

Markov chain usage model is a probabilistic model based on FSM (Finite State Machine) model. It can be generated from textual use cases [7, 8] and then the most likely usage scenarios are identified as test cases. Since these works are based on the normal use cases, with these traditional approaches important components and events can not be identified.

Kallepalli and Tian developed UMMs (Unified Markov Models) to support statistical testing, performance evaluation,

This paper was supported by Research Fund, Kumoh National Institute of Technology

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

and reliability analysis [9]. They applied the model to web application testing and proposed an automated way of constructing UMMs model from web server logs [9, 10]. With this approach, one web page or one information file serves as a state, so components such as java components and events cannot be included in the scenarios. Due to the complexity of the model level in the web applications, identifying important components and events did not be covered.

Daniel Amyot applied UCM model to customer-oriented acceptance tests for Web applications [5]. But complete coverage of acceptance tests in this approach is impractical for the large projects. It is necessary to identify highly frequently used usage scenarios and components in order to achieve efficient tests to assure and maximize the product reliability.

III. BACKGROUND

A. Use Case Map

UCM is part of user requirements notation standards proposed to the ITU-T for describing functional requirements as causal scenarios. UCM allows developers to model dynamic behavior of systems where scenarios and structures may change at run-time. For these reasons, UCM has been widely used in a range of systems [3].

UCM notation consists of three formal elements that form the basis for all maps: paths represent scenarios; components including system and non-system entities perform responsibilities; and responsibilities such as actions, events and so on are linked to paths and may be contained within components to indicate that the component executes that event. UCM scenarios begin with start points which represent triggering events and/or pre-conditions which are required for the commencement of the scenario. Scenario paths are followed to endpoints that represent terminating events or post conditions of the scenario execution. The UCM notation also provides a hierarchical abstraction mechanism in the form of stubs (diamonds) and plug-ins (sub-maps). Each hierarchy of maps has a root map that contains stubs where lower-level maps can be plugged in.

UCM are primarily visual, but a formal textual representation also exists. Based on the XML (eXtended Markup Language) 1.0 standard, this representation allows for tools to generate UCMs or use them for further processing and analysis [11].

We mentioned Figure 1 to demonstrate the fundamental understanding of UCMS and it will be used as an example in the rest of the paper. The figure shows a simple UCMs model where UserO attempts to establish a telephone call with another UserT through some network of agents. Each user has an agent responsible for managing subscribed telephony features. UserO first sends a connection request (req) to the network through his agent. UserO's agent has an originating dynamic stub SO which has two plug-ins, default originating and OCS. UserO can subscribe to one of services. The OCS plug-in shows an object OCSlist that represents a list of screened numbers that the originating user (UserO) is forbidden to contact. The called number is checked against the list (chk). If the call is denied, a relevant message is prepared for the originating party (md). This successful request causes the called agent to verify (vrfy) whether the called party is idle or busy (terminating plug-in in stub ST). If he is idle, then there will be some status update (upd) and a ring signal will be activated on UserT's side (ring), concurrently, preparation of a ring-back targeted to the originating party (mrb). Otherwise, a message stating that UserT is not available will be prepared (mb) and sent back to UserO (msg). For a detailed description of UCMs notation, readers should refer to [2] and [12].

Compared to the traditional use case diagrams, UCMs contain more object information such as components, responsibilities, stubs and plug-ins. Thus this paper applies statistical technique to UCMs model which can analyze not only scenarios but also these additional objects to assure and maximize the product reliability from a customer's perspective.

B. Markov chain usage models

Markov chains is the simplification based on the so called memoryless property, or the Markovian property, which states that the state transitions from a given state depend only on the current state, but not the history or how we reached that particular state [13].

As shown in Figure 2, a usage chain consists of states, The Markov chain is completely defined when transition probabilities are established that represent the best estimate of real usage. For each state the summation of output arcs or probabilities should be 1. Transition probabilities can be obtained from various sources containing information about the actual usage counts and relative frequencies. Several methods can be employed to extract this information, including subjective evaluation based on expert opinions, survey of target customers, and measurement of actual usage logs.

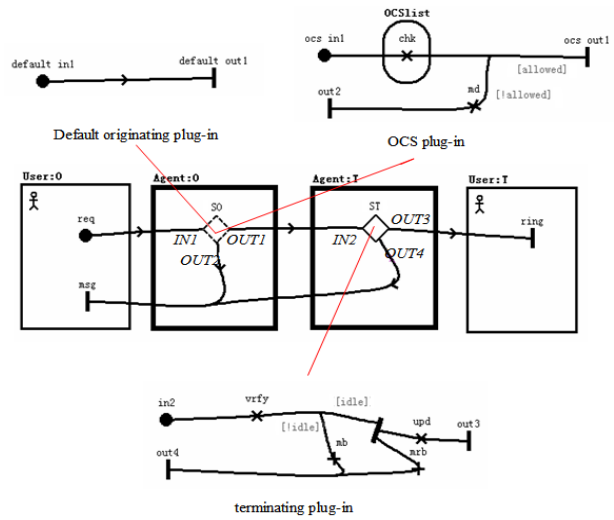


Figure 1. A Simple telephone system UCMs model

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

An operational sequence consisting of visits to multiple states can be constructed by following the state transitions. The likelihood for a particular sequence to happen can also be easily calculated by the product of its individual state transition probabilities. Therefore, Markov chains can be used to ensure performance and reliability based on usage scenarios and frequencies by target customers. Furthermore, Kallepalli and Tian developed UMMs to support statistical testing, performance evaluation, and reliability analysis. UMMs possess a hierarchical structure formed by a collection of Markov chains. It captures information about execution flow (control flow), transaction processing (workload creation, handling, and termination), and associated probabilistic usage information.

As introduced in section 3.1, UCMs model also has a hierarchical structure, a root map that contains stubs where lower-level maps can be plugged in. Thus it is straightforward to convert UCMs model to UMMs.

For example as shown in Figure 3, the top level Markov chain, depicted by the upper part of the diagram, represents high-level operational units (states), associated connections (transitions), and usage probabilities. Various sub operations may be associated with an individual state and can be modeled by more detailed models or sub models, as shown in the down Markov chains.

IV. An Importance Analysis of UCM with Markov Chains

This paper proposes an importance analysis of UCMs model with Markov chains which can identify highly frequently used usage scenarios and generate lists of important objects such as components, responsibilities, stubs and plug-ins in UCMs specifications. This section will introduce overall phases and detail procedure with an example.

The main procedure of this approach is illustrated in Figure 4.

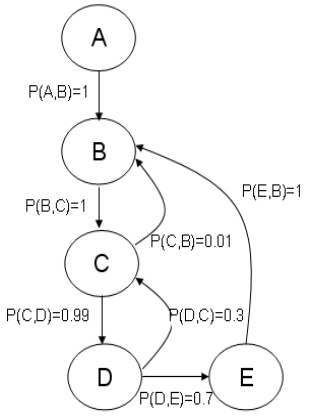


Figure 2. A Markov usage chain example

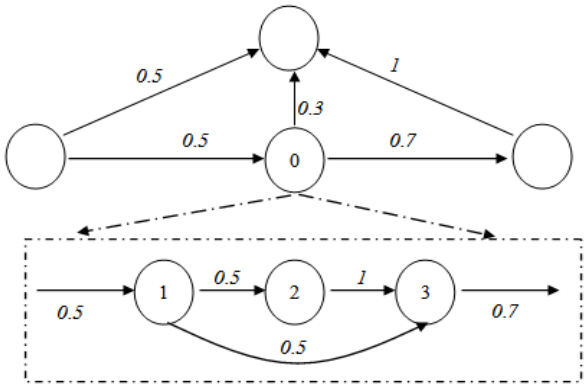


Figure 3. UMMs example

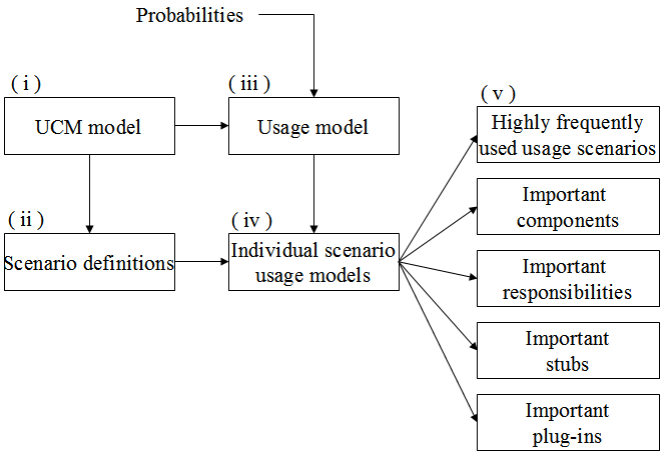


Figure 4. Overall procedure of the proposed approach

This approach includes five steps as follows:

1) Construct a UCMs model during requirement phase: A UCMs model should be constructed during requirement phase (i); UCMNav or jUCMNav tool can be used to draw UCM notations [3].

2) Define all the possible scenarios from the UCMs model: Since it is easier to identify transition probability on scenario basis; the next step is generating all of the possible scenarios from the UCMs model. The generation of scenarios is for Transition probabilities should be prepared for each scenarios path

3) Convert the UCMs model to an usage model with probabilities: After assigning probabilities, the UCMs model is converted to a usage model based on some rules that will be discussed later

4) Generate each individual scenario usage model.

5) An importance analysis of the objects of each object types of the UCMs: With the individual scenario usage models, an importance analysis should be followed to identify highly frequently used usage scenarios and generate lists of objects (components, responsibilities, stubs and plug-ins) ranked by their relative importance factor in the early phases of software life cycle.

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

TABLE I. SCENARIO DEFINITIONS FOR SIMPLE TELEPHONE SYSTEM

Scenario name	Pre-condition	Post-condition
NormalIdleCall	Start point = req SO = default in1 ST = in2 [idle] = true	null
NormalBusyCall	Start point = req SO = default in1 ST = in2 [idle] = false	null
OCSDeniedCall	Start point = req SO = ocs in1 [idle] = false	null
OCSAllowedIdleCall	Start point = req SO = ocs in1 ST = in2 [allowed] = true [idle] = true	null
OCSAllowedBusyCall	Start point = req SO = ocs in1 ST = in2 [allowed] = true [idle] = false	null

TABLE II. THE CORRESPONDENCES BETWEEN UCMS MODEL AND USAGE MODEL

UCMs model	Usage Model
Start/end points	States
Responsibilities	States
Scenario path	Transition path
AND/OR forks and joins	AND/OR forks and joins

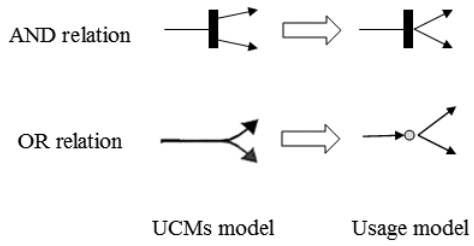


Figure 5. "AND/OR" relation translation

B. An UCMs model construction

UCMs graphical models describe functional requirements and high-level designs with causally linked responsibilities, superimposed on structures of components. With the help of editor UCMNav or jUCMNav, it is easy to construct UCMs model. A detailed introduction of how to use these editors can be found in [3].

For an example, a UCMs model of a simple telephone system is illustrated in Figure 1. It is designed with jUCMNav editor.

C. Identifying Scenario definitions for UCMs model

For extracting individual scenarios, all possible scenarios should be defined. For each scenario, one needs to provide a scenario name, pre-condition including the list of start points to be triggered and initial values of global Boolean variables, and (optionally) a post-condition used to assert the validity of a scenario once the traversal has completed.

As shown in Figure 1, a simple telephone system is modeled with jUCMNav editor. The following five scenarios are identified from the UCMs model as shown in Table 1:

1) NormalIdleCall: UserO subscribes default originating plug-in and gives a call to UserT who is idle.

2) NormalBusyCall: UserO subscribes default originating plug-in and gives a call to UserT who is busy.

3) OCSDeniedCall: UserO subscribes originating call screening plug-in and gives a call to UserT, but this number is in the list of screened numbers that the originating UserO is forbidden to contact.

4) OCSAllowedIdleCall: UserO subscribes originating call screening plug-in and gives a call to UserT whose number is allowed and UserT is idle.

5) OCSAllowedBusyCall: UserO subscribes originating call screening plug-in and gives a call to UserT whose number is allowed and UserT is busy.

D. Converting UCMs model to usage model with probabilities

The general approach of usage model construction includes two steps. First, construct a basic model with basic states (nodes) and state transitions (links) identified from product specification. Second, complete the usage model by assigning transition probabilities.

There are some quite close correspondences between some of the scenario entities in UCMs model and usage model elements. Table 2 illustrates these correspondences.In the Table 2, UCM start/end points or responsibilities can represent states in usage model. Causal relationships path between responsibilities in UCMs model can serve as transition path between states in usage model. AND/OR forks and joins relationships serve as AND/OR forks and joins in usage model. The AND/OR relation is shown in Figure 5.

"Stub" elements in UCM root map can be converted to states in top-level models of UMMs. "Plug-in" elements in UCM sub-map can be converted to sub-level models of UMMs as shown in Figure 6. The big round with "S" means a state. The "P1" and "P2" identify the path1 and path2. If the "Stub" is a dynamic stub, an additional "or" relation by a small round notation is added.

The second step of the construction of a usage model is assigning transition probabilities. Several methods such as expert opinions, survey of target customers, and measurement of actual usage logs can be employed to extract these probabilities [10]. Basic UCMs model is augmented with

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

assigning probabilities to start points, including plug-ins' start points, and branches on OR-forks.

As shown in Figure 1, a simple telephone system has been modeled by UCM. This UCMs model should be converted to usage model that is illustrated in Figure 7. This usage model is an UMM, the top-level UMM represents the root map in the UCMs model, and there are 2 sub-level UMMs which represent plug-ins of two stubs in top-level model. The numbers assigned to transitions represent transition probabilities.

E. *Individual scenario usage model generation*

As shown in Table 1, five scenarios are defined for the example simple telephone system. Thus based on the whole usage model, individual scenario usage models can be generated. Figure 8 shows NormalIdleCall scenario usage model. Other four scenarios usage models can also be generated with the same way.

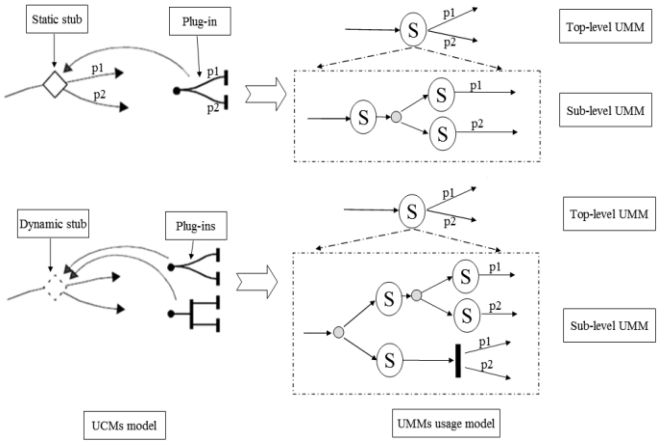


Figure 6. "stub" and "plug-in" elements of UCMs in UMMs

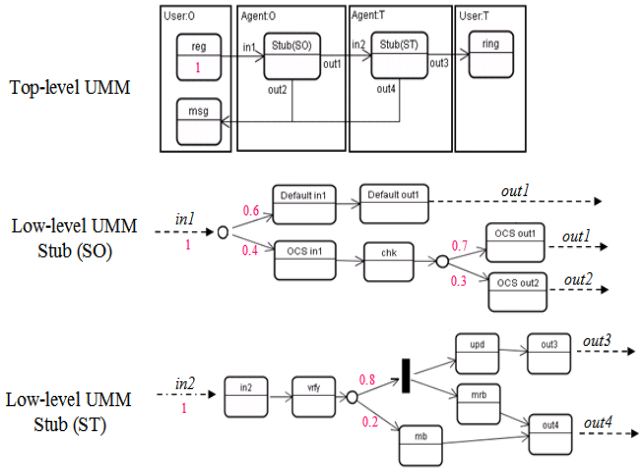


Figure 7. Usage model for the simple telephone system

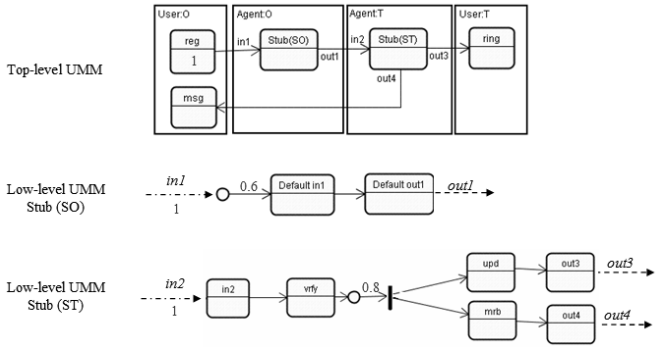


Figure 8. NormalIdleCall scenario usage model

F. *An Importance Analysis of UCMs*

In this section, an importance analysis methodology of UCMs model will be introduced. Firstly we will provide several definition of the model. Secondly with an example the insight of the equations for importance will be illustrated. Finally the equations are presented.

The definition of importance of an object is given as Definition 1. According to the definition, if an object is used more frequently by users, the importance of the object is higher.

Definition 1: *Importance of an object is usage probability of the object from a customer's perspective.*

As shown in the background of UCMs, there are several kinds of objects existed in UCMs model such as scenarios, components, responsibilities (start/end points), stubs and plug-ins. According to the description of UCMs, a component can contain responsibilities, stubs, plug-ins, and other components. And also a plug-in can contain responsibilities, stubs and components. We can easily notice that is has a composite pattern. Therefore all the objects in the UCMs model are classified as three types: scenarios, primitive objects and containers. The definitions of these objects are given as follows:

Definition 2: *A scenario is a path which consists of transitions from start point to end point across the related responsibilities which may bind to the components*.

Definition 3: *A primitive object is an object which is not includes any other objects*.

Definition 4: *A container is an object that consists of other objects such as primitive objects* or containers.

The importance of a scenario can be calculated as product of related individual transition importance or probability. For a primitive object, the importance is summation of product of each related scenario's importance and the number of appearances of primitive object in this scenario. The importance of a container can be calculated as summation of the importance of its contained objects.

Figure 9 shows an example of an object model to illustrate the insight of the calculation of the importance of each type of the object. This object model is actually a lattice model. In this example, four scenarios (S_1, S_2, S_3 and S_4) are defined, and

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

there are six primitive objects and four containers. Scenario S_1 has three transitions T_1, T_3 and T_5. Because the transition probabilities should be prepared initially, the importance of scenarios S1 can be calculated as $I(S_1) = I(T_1) * I(T_3) * I(T_5)$. The primitive object R01 appears in scenario S_1 once and in scenario S_2 also once. So the importance of primitive object R01 can be calculated as $I(R01) = I(S_1) * 1 + I(S_2) * 1$. For the container C03, it contains two child objects R01 and R03. Thus the importance of the container C03 is calculated as I (C03) = I (R01) + I (R03). By this way, the importance of all the objects can be calculated.

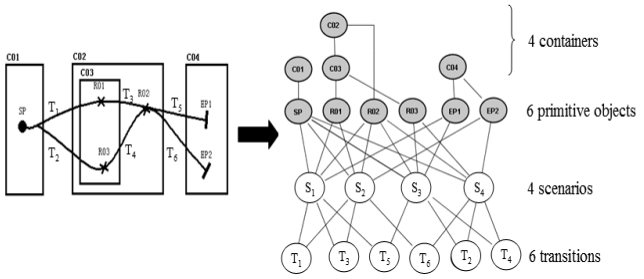


Figure 9. An example of object model for importance analysis of UCM

After generalizing the above example we can define formal equations of calculating the importance of all the objects in UCMs model as follows. The Eq.1 is used for the importance of a scenario. The Eq.2 is for the importance of a primitive object and the Eq.3 is used for a container.

$$I(S) = \prod_{T_i \in S} I(T_i) \qquad (1)$$

Where:

S represents a scenario

I(S) is importance of a scenario S

T_i is a transition included in scenario S

$I(T_i)$ is importance of a transition T_i in this scenario which is initially defined

$$I(PO) = \sum_{S_i \in S(C)} \left(I(S_i) \times N(PO, S_i) \right) \qquad (2)$$

Where:

PO represents a primitive object

I(PO) is importance of a primitive object PO

S_i represents a scenario

$I(S_i)$ is importance of scenario S_i

S(C) is a set of scenarios which includes the primitive object PO

N(PO, S_i) is the number of the primitive object PO appeared in scenario S_i

$$I(C) = \sum_{O_i \in O} I(O_i) \qquad (3)$$

TABLE III. SCENARIOS IMPORTANCE FOR THE SIMPLE TELEPHONE SYSTEM

Scenarios	Importance
NormalIdleCall	0.48
NormalBusyCall	0.12
OCSDeniedCall	0.12
OCSAllowedIdleCall	0.224
OCSAllowedBusyCall	0.056

TABLE IV. IMPORTANT SCENARIOS WITH OVERALL IMPORTANCE THRESHOLD 0.2

Scenarios	Importance
NormalIdleCall	0.48
OCSAllowedIdleCall	0.224

TABLE V. TABLE 5. IMPORTANT SCENARIOS WITH ALTERNATIVE IMPORTANCE THRESHOLD 0.3

Scenarios	Importance
NormalIdleCall	0.6*0.8=0.48
OCSDeniedCall	0.4*0.3=0.12
OCSAllowedIdleCall	0.4*0.7*0.8=0.224

Where:

C represents a container

I(C) is importance of a container C

O(C) is a set of child objects of container C.

O_i represents a contained object in the container C. The contained objects may be child primitive objects or containers

$I(O_i)$ is importance of object O_i

G. Case study of the importance analysis of UCM: a simple telephone system

In this sub-section we will show an example according to the importance analysis method explained in 4.5. For the simple telephone system, five scenarios are identified from the UCMs model (Table 1). Thus by Eq. (1), the importance of these five scenarios is calculated and the result is shown in Table 3.

After calculating importance of scenarios, to identify the highly frequently used usage scenarios, some thresholds are used. In practical applications, thresholds can be adjusted to control the numbers of test cases to be generated and executed. Several thresholds have been initially proposed in [14] and used in developing UMMs. Two kinds of thresholds are used in this paper:

1) Overall importance threshold for complete end-to-end operations to ensure that commonly used complete operation sequences by target customers are covered and adequately tested.

2) Alternative importance threshold to ensure commonly used operation pairs, their interconnections and interfaces are covered and adequately tested.

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

TABLE VI. IMPORTANCE OF ALL RESPONSIBILITIES IN SIMPLE TELEPHONE SYSTEM

Responsibilities	Importance
req	1
msg	1
vrfy	0.88
out4	0.88
in2	0.88
upd	0.704
ring	0.704
out3	0.704
mrb	0.704
default out1	0.6
default in1	0.6
ocs in1	0.4
chk	0.4
ocs out1	0.28
mb	0.176
out2	0.12
md	0.12

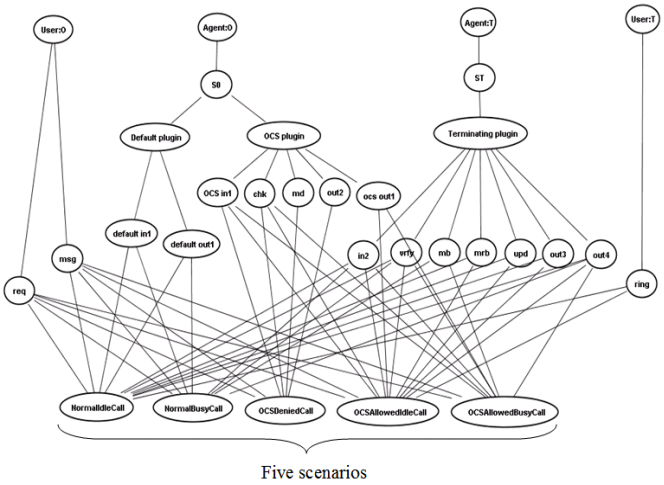


Figure 10. Composition Object model of UCMs model for simple telephone system

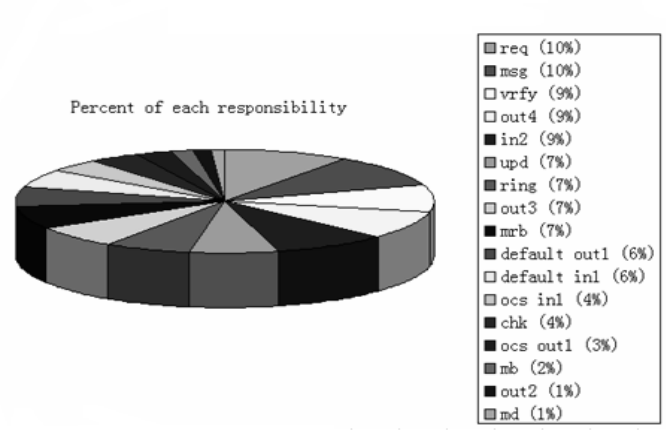


Figure 11. Percent of each responsibility importance

So if overall importance threshold is given as 0.2, then the highly frequently used usage scenarios can be identified as shown in Table 4. If alternative importance threshold is given as 0.3, Table 5 shows the highly frequently used usage scenarios.

For calculating the importance of components, responsibilities, stubs and plug-ins in UCMs model of simple telephone system, firstly a composition object model as shown in Figure 10 is given. This object model clearly describes the architecture of relations among different objects in UCMs model. Based on this object model, importance of each object can be easily calculated.

Responsibilities in this object model are all primitive objects, so with Eq. (2), the importance of all the responsibilities can be calculated as shown in Table 6.

Plug-ins in this object model are all containers, importance of each plug-in can be calculated as summation of its children's importance by Eq. (3). The result is shown as follows:

- Importance of plug-in Default ORIGINATING = 1.2
- Importance of plug-in TERMINATING = 4.928
- Importance of plug-in OCS = 1.32

Importance of each stub and component also can be calculated with Eq. (3). The result is shown as follows:

- Importance of stub SO = 2.52
- Importance of stub ST = 4.928
- Importance of component AgentT = 4.928
- Importance of component AgentO = 4.928
- Importance of component UserO = 2.52
- Importance of component UserT = 0.704

After getting all objects importance, a percent of importance for each object type can be calculated to show which objects are more important clearly. Percent of responsibility importance is shown in Figure 11.

The calculation of the importance is very simple and straightforward. As presented above the importance of each scenario can be easily calculated by Eq. (1) using transition probabilities, and after getting the importance of scenarios, the importance of primitive objects can be calculated with Eq. (2) based on the object model. For containers, the importance can be calculated as simple summation of its children's importance by Eq. (3).

In the early phases of the software life cycle (requirement analysis phase) the importance of all scenarios, components, responsibilities, stubs and plug-ins, and the highly frequently used usage scenarios and important objects can be identified from a customer's perspective. Since the importance of each object type can be identified at the early phase where UCM is defined, the quality engineering efforts can be applied from

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

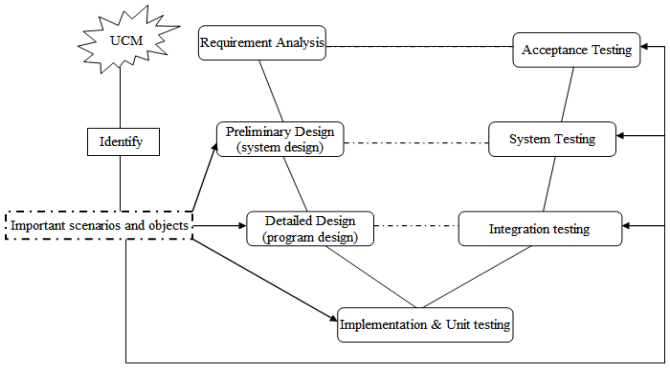


Figure 12. UCM in V model

first phase to the last phase. That means this approach can support in the full software life cycle for quality assurance activities. For instance as shown in Figure 12, since the importance of each objects is identified at the Requirement Analysis phase, the important scenarios can be heavily tested on the Acceptance test, the important containers are can be took cared at the system testing and integration testing and the important primitive object can be focused at unit testing.

V. CONCLUSION

This paper proposes an importance analysis of UCM with Markov chains which applies statistical technique to UCMs model. Like other traditional statistical techniques, highly frequently used usage scenarios can be identified.

Furthermore, since the approach of this paper is based on UCMs model, it can identify more important object types such as components, responsibilities, stubs and plug-ins. Therefore this paper also proposes an object model of UCM which can calculate the importance of these objects. Thus the efficient testing can be possible for both highly frequently used usage scenarios and important object types to assure and maximize the product reliability from a customer's perspective. Since the combination of UCMs model and statistical technique can identify much more important information, it is very beneficial for quality assurance engineering.

The importance of each object is identified at the Requirement Analysis phase, which means the approach make possible the quality driven development during the full software life cycle. The method of the importance calculation is very clear, easy and straightforward that it is very practical to the real environment. To show these characters we include a simple telephone example with importance calculations.

In this approach, it requires manual operation to generate usage model and object models. Future work will extend the UCM Navigator (UCMNav or jUCMNav) editor to support automatic generation. Future work will also include the research of an algorithm to generate all possible scenarios automatically.

REFERENCES

[1] R.J.A. Buhr, "Use Case Maps as Architectural Entities for Complex Systems", IEEE Transactions on Software Engineering, Special Issue on Scenario Management. December 1998, Vol. 24, No. 12, pp. 1131-1155.

[2] "Use Case Maps Web Page and UCM User Group", http://www.UseCaseMaps.org, 1999.

[3] J.A. Whittaker, and M.G. Thomason, "A Markov chain model for statistical software testing", IEEE Trans on Software Engineering, Oct. 1994,Vol. 20, No. 10, pp. 812–824.

[4] Amyot, D., Specification and Validation of Telecommunications System wtih Use Case Maps and *Lotos*, in *School of Technology and Engineering*. 2001, University of Ottawa: Ottawa.d

[5] D.Amyot, J.F. Roy, and M.Weiss, "UCM-Driven Testing of Web Applications", 12th SDL Forum, June 2005, pp. 247-264.

[6] D. Amyot, X.Y. He, Y. He and D.Y. Cho, "Generating scenarios from use case map specifications", Third International Conference On Quality Software (QSIC), Dallas, USA, Nov. 2003, pp. 108–115.

[7] M. HÜBNER, I. PHILIPPOW, and M. RIEBISCH, "Statistical Usage Testing Based on UML", Proceedings of the 13th Annual IEEE International Symposium and Workshop on Engineering of Computer Based Systems, 2006.

[8] B. Regnell, P. Runeson, and C. Wohlin, "Towards Integration of Use Case Modelling and Usage-Based Testing", Journal of Systems and Software, 2000, pp. 50:117-130.

[9] C. Kallepalli, and J. Tian, "Measuring and Modeling Usage and Reliability for Statistical Web Testing", IEEE Transactions on Software Engineering, Nov. 2001, vol. 27, no. 11, pp. 1023-1036.

[10] C. Kallepalli, and J. Tian, "Usage Measurement for Statistical Web Testing and Reliability Analysis", Proceeding of Seventh International Software Metrics Symposium, London, IEEE Computer Society Press, 2001, pp. 148-158.

[11] D. Amyot, and A. Miga, Use Case Maps Linear Form in XML, version 0.13, May 1999.

[12] R.J.A. Buhr, and R.S. Casselman, "Use Case Maps for Object-Oriented Systems", Prentice-Hall, USA, 1995.

[13] S. Karlin, and H. M. Taylor, A First Course in Stochastic Processes, 2nd Ed. Academic Press, New York, 1975.

[14] J. D. Musa. Software Reliability Engineering. McGraw-Hill, New York, 1998.

AUTHORS PROFILE

Yaping Feng received his B.S. degree in softtware engineering from East China Normal University, China in 2005 and received his M.S degree in computer engineering from Kumoh National Institute of Technology, Korea in 2007. He has worked as a developer in POSData China from 2008. His current research interests include Software Engineering and Mobile Computing.

Lee-Sub Lee received his B.S., M.S. degrees in mathematics and computer science from Sogang University, Seoul, Korea, in 1988 and 1990, respectively. He ha a Ph.D degree in computer science & engineering from Korea University, Seoul, Korea in 2004. He is assistant professor of Kumoh National Institute of Technology since 2004. He had worked as a senior manager of the IT R&D Center, Samsung SDS, Ltd from 1990 in SungNam, Korea. His current research interests include Software Engineering and Mobile Computing.

Convergence of Corporate and Information Security

Syed (Shawon) M. Rahman, PhD
Assistant Professor of Computer Science
University of Hawaii-Hilo, HI, USA and
Adjunct Faculty, Capella University, MN, USA

Shannon E. Donahue CISM,CISSP
Ph. D. Student, Capella University,
225 South 6th Street, 9th Floor
Minneapolis, MN 55402, USA

***Abstract*—As physical and information security boundaries have become increasingly blurry many organizations are experiencing challenges with how to effectively and efficiently manage security within the corporate. There is no current standard or best practice offered by the security community regarding convergence; however many organizations such as the Alliance for Enterprise Security Risk Management (AESRM) offer some excellent suggestions for integrating a converged security program. This paper reports on how organizations have traditionally managed asset protection, why that is changing and how to establish convergence to optimize security's value to the business within an enterprise.**

Keywords-component; convergence; security; risk management; corporate; threats

I. INTRODUCTION

Throughout history organizations have had security and loss prevention departments to protect their physical assets. In the last 10 to 20 years however there has been a major shift in what is considered an asset. Information and intangible assets have increased significantly. In fact, one could make a very strong argument that information has become an enterprise's most important asset. In the not so distant past, organizations considered their most valuable assets to be physical assets. With the growth of the internet and increased methods for communication which has given most organizations the ability to do business globally, along with the expansion of data warehousing and electronic storage, information has fast become the most critical element to the success of an organization [1].

Information provides organizations with data on their customers, finances, inventories, suppliers, partners and competitors. Metadata (data formed by combining groups of information) provides organizations with vital information that organizations use for decision making on a daily basis. Even the organizations who do still mainly rely on physical resources need information to forecast and communicate with vendors and suppliers.

Before intangible assets became the largest value to organizations throughout the world, most companies counted their physical assets as their primary asset. Organizations had loss prevention units and security departments that safeguarded the company's assets with cameras, physical access control measures, and security guards.

These corporate security departments were made up largely of former law enforcement officers that reported to legal, compliance or risk management divisions. The main charges of this department were intrusion protection and investigations in the workplace. As the digital age came to life, organizations saw more and more information being stored on servers, in databases, and in files.

Organizations began to realize the value that intangible assets provide and subsequently created departments responsible for securing that information. IT security departments and information security departments were tasked with managing the risk that surrounded information.

These information security departments were created after organizations saw threats to their information in the form of hackers, malware, unavailability, and data theft. Additionally, both regional and international laws have surfaced which require dedicated information security managers to be responsible for a formal information security program which is responsible for data protection. Many of the people responsible for information security were moved into the role due in part to their technical backgrounds which were very useful when considering technical controls and designs of a logical perimeter. However, due to their technical backgrounds most of these information security professionals had no experience in traditional corporate or physical security.

Initially having two separate groups responsible for the security of different types of assets was not a problem. However, some security functions have begun to converge on their own and this creates unique challenges for organizations with stove piped security functions.

As technology evolved corporate security departments began to use some advanced tools and technology such as traditional closed circuit cameras running over IP networks. The cameras were the responsibility of the corporate security team, but IT had control of the IP network. The same problem happened with access cards. The corporate security department traditionally had control over physical access control but with databases housing all of the data, IT and information security again were clearly involved. Vendors have recognized this

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

trend and have began to offer converged solutions for risks such as access control [12].

It occurred to many that having the corporate and information security departments work together may be able help to reduce an organization's overall risk profile while streamlining some redundant security processes. However, convergence did not look so appealing to everyone. While the unintentional convergence of systems was happening, the two departments who were initially affected were not working together to improve the risk.

When discussing information and corporate security, convergence is defined as "a trend affecting global enterprises that involves the identification of risks and interdependencies between business functions and processes within the enterprise and the development of managed business process solutions to address those risks and interdependencies" [1]. This definition speaks to the need for organizations to look at convergence and begin to disassemble the organizational silos in order to encourage collaboration to manage risk holistically. Fostering an environment rich in collaboration will help the organization lower its overall risk and reach its business objectives.

According to a 2007 article in security magazine [15], "the silo approach to managing enterprise risks is inadequate because it leaves too many gaps and provides no reliable way to evaluate an enterprise's risk position" [15]. Due in no small part to these gaps, many organizations have tried to begin to explore converging the functions. In reality, convergence between information and corporate security is still very immature. In most organizations, even if the departments both report to a single Chief Security Officer (CSO) or Chief Risk Officer (CRO) the departments are still as different as apples and oranges acting in their traditional silos and worse than that oftentimes shutting one another out.

II. Reasons for Convergence

One reason for convergence is that threats continually increase and become threats to both corporate and information security safeguards. If an organization's corporate and information security departments do not work together they may miss out on valuable information that could be beneficial to both areas regarding particular risks and threats. Additionally, working in silos may be detrimental as downstream risks that are an emerged result of other risks may not be considered.

United States Computer Emergency Readiness Team (US-CERT)[2] interacts with federal agencies, industry, the research community, state and local governments, and others to collect reasoned and actionable cyber security information and to identify emerging cyber security threats. Based on the information reported, US-CERT was able to identify the following cyber security trends (figure 2 and 3) for fiscal year 2009 first quarter [2].

Figure 1 displays the overall distribution of cyber security incidents and events across the six major categories. The percentage of Category 5 (Scans, Probes, or Attempted Access) reports decreased for the second consecutive quarter. This was a 2.9% decrease in CAT 5 incidents compared to the previous quarter. The percentage of Malicious Code incidents increased by 3.3%. Figure 2 is a breakdown of the top five incidents and events versus all others. Phishing remained the most prevalent incident type, accounting for 70% of all incidents reported. This was a slight percentage decrease of 1.8% from the previous quarter. on the other hand, The sophistication of attack tools has gone up, while the level of skill required to use those tools has gone down(refer to figure 3). At this stage, the attacker takes advantage of his or her ability to steal confidential and proprietary data and sells it for profit or uses it for military intelligence [3].

Another strong argument for convergence is the blurring of boundaries between corporate and information security. For example, if a corporate security department is in charge of corporate access control to restricted areas via card readers, but IT owns the systems who responds when there is a major breach? Likewise, if there is a disaster and information security and corporate security are not aligned with their plans whose plan does the organization follow ?

Convergence offers the organization the opportunity to restructure systems. Currently, systems in physical and information security are oftentimes segregated from one another and are not aware of what the other systems are doing. Once an organization decides to move forward with convergence, systems can be combined which saves the

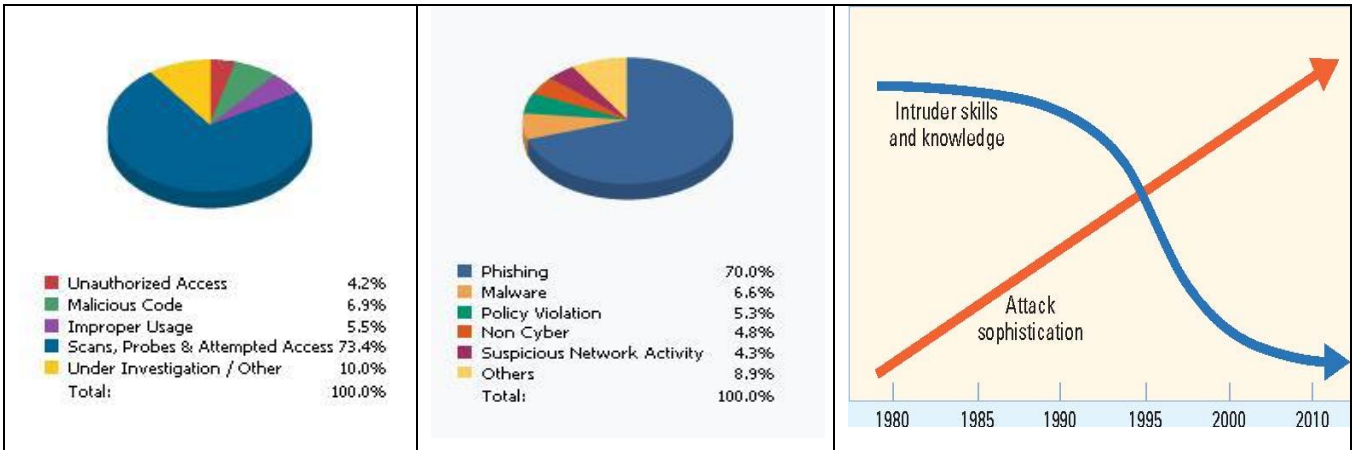


Figure 1: Incidents and Events by Category[2]

Figure 2: Top Five Incidents vs. All Others [2]

Figure 3. Attack tool trends[3]

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

organization money on upkeep and maintenance of infrastructure, not having to purchase any new hardware and also lowers bandwidth uses on the company network.

III. BENEFITS OF CONVERGENCE

Joining the information and corporate security departments in some way will help to dispel some of the device management confusion. Whether it is through just working together to have information security and corporate security handle a joint presentation to promote awareness training, or actually joining the two separate functions to report to one CRO or one CSO benefits can be derived from some sort of convergence. If organizations choose to organizationally link the two functions there will be some immediate cost savings as a Chief managing both areas should have an understanding of which team members should be managing particular areas and can set goals that are achievable by both departments. This helps to enable the departments to work together towards a common goal and hopefully will reduce the amount of overall risk to the enterprise.

Gaps between corporate and information security have caused problems in the past and convergence including collaboration and training can help to minimize these gaps. For example, if the information security department has all controls applied appropriately and theft occurs as a result of a thief posing as an employee the breach has still occurred. In fact, a good example of information and corporate security not being on the same page occurred recently at the Sumitomo Mitsui Bank in London, England. Criminals who posed as janitors within the bank had installed devices on computer keyboards that allowed them to obtain login information. The criminals tried unsuccessfully to steal £220 million. Information security controls had been applied but there was a physical security breach that could have been devastating to the organization [6].

While convergence could surely have helped in the above mentioned situation, many organizations are slow to adopt a converged approach. CSO's and CISO's have spoken up and want the business community to understand the benefits of convergence. It appears that business leaders are becoming more accepting of the idea of convergence. According to a 2005 PricewaterhouseCoopers and *CIO* magazine survey of 8,200 IT and security executives in 63 countries of, 53% of organizations have some level of integration between their corporate and IT security divisions [8]. That's up from just 29% in 2003. The projections seem to be growing as well. The Alliance for Enterprise Security Risk Management expected 2005 global spending on convergence activities expected to reach $1.1 billion dollars and significantly move upward after that.

Some of the major benefits that organizations are seeing as a result of converged information and corporate security are significant. Many organizations are saving millions of dollars by streamlining these functions.

Other benefits exist as well. It is hard to put a dollar amount on how safe people feel in the office, or how not being the latest company to lose millions of customer's data is effecting the organization. But through these types if situations, security is not only keeping its assets safe but is preventing unnecessary funds from going out the door as well.

According to a computerweekly.com report executives are seeing the benefits of convergence to organizational risk, "According to the results of a global survey conducted by the Economist Intelligence Unit (EIU) for AT&T, the majority of executives (52%) believe that having a converged network gives their companies a better defense against IT security breaches" [11].

TABLE 1: GLOBAL SECURITY CONVERGENCE SPENDING FORECAST [4].

	2004	2005	2006	2007	2008
Large-scale convergence projects in NA and Europe	19	68	175	382	856
Physical/logical access control projects in NA and Europe	50	150	413	903	1,656
Other projects performed jointly by IT and physical security departments in NA and Europe	13	45	118	246	406
Public sector: border control convergence systems, law enforcement projects in NA and Europe	410	820	1,899	4,202	8,003
Small projects (data center security, communications security, etc.) in NA and Europe	14	40	108	229	369
Total	506	1,123	2,713	5,962	11,289

Cost control and productivity may also be improved as a result of convergence. By having systems joined, an organization can eliminate steps that are redundant. This helps to improve processes, eliminate human errors, and increase productivity which ultimately generates revenue for the organization.

Organizations have learned quite a bit about business continuity and disaster recovery in the last few years. September 11, Hurricane Katrina, and devastating natural disasters around the world have brought an increased focus on business continuity (BCP) and disaster recovery planning (DRP). BCP and DRP are another area where convergence plays an important role. We believe information technology security and corporate security must work together to ensure that they are meeting one another's business needs when it comes to project their assets from inside and outside threats and recover data and resources from any attacks.

The collaboration will also streamline efforts so that everyone can understand what to do in cases of emergencies. Collaborating on these efforts will be instrumental in restoring

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

services in a timely manner. Restoration of these services is an extremely critical issue and is an indirect revenue generator since the quicker that services are restored the faster money is being made.

While the benefits of convergence seem very obvious, there are quite a few challenges that will create some issues for the person who is charged with convergence.

IV. CHALLENGES OF CONVERGENCE

Security convergence is not without its challenges. The information security department and corporate security departments have long operated in parallel. They have not shared and have not wanted to and, in a lot of situations have through of the other as an adversary. In order to have convergence be effective, collaboration between these two areas must happen. It is not enough to have them reporting through the same channel, they must both have common goals and have mutual discussions on how to meet those goals. In order to get to this level of cooperation there are huge cultural challenges that will need attention.

Information security professionals and corporate security professionals do not usually have the same background. Many information security professionals have technical backgrounds while many of the corporate security professionals have law enforcement backgrounds. According to Steve Hunt, President of 4A international LLC a security consulting firm in Chicago, these differences can lead to a gap in how the two departments evaluate security technologies and controls [8].

Salaries are another issue. Corporate security professionals are not earning comparable salaries generated by information security professionals. Information security professionals often collect six figure salaries early in their career whereas the traditional corporate security professional might be making only half of that closer to the end of their career. Parity adjustments are not practical because there are ways to justify the salary disparity such as technical skills, higher levels of education, professional certifications, and a better job market. However, when an organization combines groups and has them working in tandem, if half of them are earning 50% of what the others are earning there will be some bitter feelings which may result in a hostile environment where people will not work together.

One additional challenge is the training gap. Corporate security professionals often have not been trained on information security or technology. They are not aware of what to watch for when performing their own duties. Often they are unhappy with the convergence plan because they are concerned that they will not retain their positions. A cross training plan that would help the corporate security people understand better what social engineering is and how to spot it, what phishing is and how to spot it and some other technical training would go a long way towards team building and improved skill sets. Additionally, information security professionals likely have very little skill in surveillance, disaster response, investigations and loss prevention. Both sets of individuals need to have some basic training so they are at least aware of what the other is doing and how it complements their own security efforts.

Organizational culture is the most difficult thing to influence so any help from senior management is going to have added benefit. It is important that senior management really demonstrate support for the security program so that employees understand that it is a critical piece of the business that is everyone's responsibility in the same way that customer service and quality management are.

TABLE 2: SUMMARY OF BENEFITS AND CHALLENGES OF CONVERGENCE

Benefits and challenges of Convergence	
Benefits	**Challenges**
Cost Savings	Culture
More holistic view of risk	Salary Differences
Reduction of risk profile	Training requirements
Streamline process	Lack of Collaboration

V. BEGINNING A CONVERGED PROGRAM

Once the organization decides that they want to converge their security program there are many things to consider. Organizational structure is one, budgets are another and the overall risk profile is a third.

A. *Organizational Structure*

Organizationally there are many routes that can be taken to align the physical and information security organizations. Each of the options has positives and negatives and may work in one organization and not at all in another. Security managers need to know the organization that they work in, consider their budgeting options as well as objectives and select a solution that will meet the needs of their business.

The first method is to combine the physical and information security departments into one security organization reporting to a CSO or CRO. This group would have responsibility for all things security from the guard at the front desk to encryption protocols for sensitive data.

While the model of combining all the players does force some level of integration, and if successful can be extremely effective in business process integration and fraud detection it is not without drawbacks. Going right to a fully immersed security program will increase the likelihood that the security staff will be upset. This unhappiness can result in a lack of cooperation and hostile work environments and may even cause damage to the risk profile because staff is so worried about their work situation that they aren't paying attention to the organizational risks.

If this is the route chosen the security manager will need to spend plenty of time working on opening communication channels and fostering collaboration. Cross training will be an immediate need if expectations are for people to learn each

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

other job functions and additional funding for corporate training may be necessary depending on the gap in skills.

The next option is to keep the physical and logical departments as separate departments with their own budgets and reporting lines but ultimately have them report to the same executive, most likely a CSO. The CSO would receive information from both security areas and be able to make decisions based on the information provided to them by the business unit director. The groups would most likely not work joint projects or go through any cross training, but would still benefit from updated processes based on organizational goals. The departments would still need to collaborate to ensure all risks are addressed. While this level of integration cannot be considered complete immersion, security does not have to be a completely combined area to reap some of the benefits of convergence. This type of management can still be extremely effective at driving down costs and eliminating redundant work and systems.

The third option which is becoming more popular these days is to keep the functions completely separate and assist in process management and collaboration through bringing security issues to a risk council staffed with business and security management that could make decisions regarding security. This approach would be helpful in minimizing culture issues and would probably be well received by members of both security teams.

Whichever organizational structure choice is taken it is important to consider the culture of the organization. Culture is an often overlooked area that can be critical to the success of implementing convergence and should not be underestimated.

Many things can be done to smooth the progress of convergence. Combining processes where possible, gaining senior management support for the initiative and beginning to look at organizational risks instead of just risks to individual departments will help to incorporate convergence smoothly.

Many frameworks exist to help an organization lower their risk profile. It is important to remember that the reason for discussing convergence of information security and physical security is to show how security professionals can help to improve the overall risk profile in your organization.

Some corporate risk management frameworks that are used internationally are the Committee of Sponsoring Organizations' (COSO) Enterprise Risk Management (ERM) framework, the Operationally Critical threats, Asset and Vulnerability Evaluation model (OCTAVE) and one more accepted framework for managing enterprise risk is the Risk and Insurance Management Society (RIMS) Risk Management Model (RMM). All of these models focus on managing operational risk and reducing the overall risk profile.

We have found the AESRM recommendation useful while beginning a converged program [4]:

- Establish a governance framework for managing security risks.
- Define security requirements early in the planning stage.
- Understand the technology better.
- Analyze and understand security-related cost-benefit trade-offs.
- Develop a unified set of meaningful standards.
- Deploy special network security controls.
- Implement effective authorization, accountability and auditability controls.
- Include critical systems in organization continuity plans.
- Protect information important for investigations.
- Increase auditing and logging.
- Require tailored training and awareness programs.
- Pressure vendors to play a more active role in security.
- Expand audit coverage of systems and devices.

VI. CONCLUSION

Organizational risks and threats are changing every day. As new technologies expand into the organization the risk profile continues to expand. Many organizations are in a constant state of growth. The organizational risk profile needs to constantly be reviewed and updated.

Since the beginning of information security departments, corporate and information security have operated in their own silos with separate management teams, different risks, different processes, and different budgets. In the last decade, organizations have seen dramatic changes in their risk profile. With technical advances, traditional corporate security functions have evolved to include the use of networks, databases, and file servers. Compliance requirements have dictated how sensitive information needs to be transmitted and so traditional corporate security departments and information security departments find some of their functions that used to have clear management delineations becoming blurry.

Although convergence has been a buzz word for a long time, companies have just recently started to notice the benefits of converging traditional siloed security functions to allow for an enterprise wide view of risk. Some of the benefits that can be derived from converging information and corporate security areas including streamlining processes, saving money on infrastructure, increasing productivity, positioning the organization better to have an broader view of organizational threats and risks and meeting compliance requirements.

Many organizations have attempted to move towards a converged security area and have realized that there are many hurdles to be jumped on the way to a converged organization. One of these challenges is getting the security staff in both areas to work together. Organizations can work to encourage collaboration by ensuring that employee concerns are addressed and people feel confident that communication will be well received. Gaining support from senior management and combining processes to minimize confusion are two more

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

techniques that will prove useful during the implementation of a converged security structure.

While the challenges may be difficult to work through, the benefits that the organization will realize as a result of taking a wider view of risk will be immense. As risks can change at any moment it is absolutely critical to have as many well trained professionals as possible working together to improve the productivity and sustainability of an organization with a holistic approach to security management through convergence in enterprise security risk management

REFERENCES

[1] Alliance for Enterprise Security Risk Management "Convergence of physical and information security in the context of risk management", 2007; http://www.aesrm.org/aesrm_convergence_in_ERM_pdf , Web retrieve on December 22, 2009

[2] US-CERT, "Cyber Security Trends, Metrics, and Security Indicators", June 16, 2009. Volume 4, Issue 1.http://www.us-cert.gov/press_room/trendsanalysisQ109.pdf, Web retrieve on December 22, 2009

[3] Liu, Simon and Cheng, Bruce; "Cyberattacks: Why, What, Who, and How", *IT Pro, IEEE Computer Society*, May/June 2009.

[4] Alliance for Enterprise Security Risk Management "Convergence of Enterprise Security Organizations", 2005; http://www.asisonline.org/newsroom/alliance.pdf, Web retrieve on December 22, 2009

[5] Alliance for Enterprise Security Risk Management "Convergent security risks in physical security systems and IT infrastructures*". ISACA Information Security Management Conference* 2006.

[6] CNN. "Police Foil $423m Cyber Robbery", Mar.2005; http://www.cnn.com/2005/WORLD/europe/03/17/bank.robbery/index.html, Web retrieve on December 22, 2009

[7] J. Burris, (2008). "Mission Impossible?", *Community Banker*, 17(6), 28-29.

[8] T. Hoffman, "Security convergence", 2006; http://www.computerworld.com/action/article.do?command=viewArticleBasic&articleId=108571&pageNumber=1, Web retrieve on December 22, 2009

[9] C. Huang, R. Behara, O. Hu, 2008; "Managing Risk Propagation in Extended Enterprise Networks"; *IT Professionals Magazine*, 10(4), 14-19.

[10] D. Radcliffe, "Companies move to combine, physical, IT security efforts" 2007; . http://www.landfield.com/isn/mailarchive/2001/Feb/0153.html; Web retrieve on December 22, 2009

[11] A. Saavas, (2007, March 22). "Converged networks deliver the best security, say execs", 2007; http://www.computerweekly.com/Articles/2007/03/22/222614/converged-networks-deliver-the-best-security-say-execs.htm, Web retrieve on December 22, 2009

[12] L. Stepanek, "Convergence here & how. SDM: Security Distributing & Marketing," 2007 . *Business Source Complete Database*, 37(2), 58-60

[13] J. Watson, Physical and IT security must go together, May 2005; http://www.computing.co.uk/computing/news/2071725/integration-way-ensure-security, Web retrieve on December 22, 2009

[14] Willoughby, M. (2003, May 28). "Bridging the gap: Information security meets physical security." http://www.computerworld.com/securitytopics/security/story/0,10801,81589,00.html, Web retrieve on December 22, 2009

[15] B. Zalud, (2007) "Neo-security: Inclusive Enterprise Risk Management. Security: For Buyers of Products, Systems & Services" *Business Source Complete Database* 44(10), 12-20.

AUTHORS PROFILE

Syed (Shawon) M. Rahman

Syed (Shawon) Rahman is an Assistant Professor in the Department of Computer Science & Engineering at the University of Hawaii-Hilo (UHH), Hawaii, USA and an Adjunct Faculty in the Capella University, Minneapolis, USA. He earned his Ph.D. in Software Engineering and MS in Computer Science degree from North Dakota State University. Dr. Rahman's research interests include Information Assurance & Security, Data Visualization, Data Modeling, Web Accessibility, and Software Testing & Quality Assurance. His research focus also includes Software Engineering Education, Search Engine Optimization, and Lightweight Software Development Methodologies such as eXtreme Programming and Test-driven Development.

Before joining at the UHH in Fall 2009, Dr. Rahman taught the last three years in the Dept. of Computer Science and Software Engineering at the University of Wisconsin-Platteville as an Assistant Professor. Dr. Rahman is always interested in applying Emerging Technologies and Tools in classrooms to improve students' learning experience and performance. He has published enormous number referred articles and presented his works around the world. He is an active member of many professional organizations including ACM, ASEE, ASQ, IEEE, and UPE.

Shannon E. Donahue

Shannon is a Certified Information Security Manager (CISM) and Certified Information Systems Security Professional (CISSP) with more than 10 years of experience in information technology and security. She was the Information Security Officer at William Rainey Harper College where she was responsible for developing and managing the information security program. She was responsible for securing data on all systems as well as incident management, policy development, and security awareness.

As a network support analyst and security architect for AT&T, Shannon provided tier 3 network support and helped to develop the virus response plan for managed service customers. She also was a lead member responsible for incident response and disaster recovery plans. In her current position with ISACA, Shannon is responsible for managing the security program and for serving the needs of the security profession through research projects and publications. Shannon has a masters degree in Management & Systems from New York University and is currently working on a PhD in Information Security at Capella University.

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

Image Retrieval Techniques based on Image Features: A State of Art approach for **CBIR**

[1]Mr. Kondekar V. H., [2]Mr. Kolkure V. S., [3]Prof.Kore S.N.
[1]Department of Electronics & Telecommunication Engineering,
([1]Walchand Institute of Technology, Solapur,
[2]Bharat Ratna Indira Gandhi Collage of Engineering, Solapur) Solapur University.
[3]Walchand College of Engineering, Sangali. Shivaji University.

Abstract-The purpose of this Paper is to describe our research on different feature extraction and matching techniques in designing a Content Based Image Retrieval (CBIR) system. Due to the enormous increase in image database sizes, as well as its vast deployment in various applications, the need for CBIR development arose. Firstly, this paper outlines a description of the primitive feature extraction techniques like: texture, colour, and shape. Once these features are extracted and used as the basis for a similarity check between images, the various matching techniques are discussed. Furthermore, the results of its performance are illustrated by a detailed example.

Keyword - CBIR, Feature Vector, Distance metrics, Similarity check, similarity matrix, Histogram, Wavelet Transform, variance, standard deviation.

I. INTRODUCTION

As processors become increasingly powerful, and memories become increasingly cheaper, the deployment of large image databases for a variety of applications have now become realisable. Databases of art works, satellite and medical imagery have been attracting more and more users in various professional fields — for example, geography, medicine, architecture, advertising, design, fashion, and publishing. Effectively and efficiently accessing desired images from large and varied image databases is now a necessity

CBIR or Content Based Image Retrieval is the retrieval of images based on visual features such as colour, texture and shape. Reasons for its development are that in many large image databases, traditional methods of image indexing have proven to be insufficient, laborious, and extremely time consuming. These old methods of image indexing, ranging from storing an image in the database and associating it with a keyword or number, to associating it with a categorized description, have become obsolete. This is not CBIR. In CBIR, each image that is stored in the database has its features extracted and compared to the features of the query image. It involves two steps:

- Feature Extraction: The first step in the process is extracting image features to a distinguishable extent.
- Matching: The second step involves matching these features to yield a result that is visually similar.

Examples of CBIR applications are:

- Security Check: Finger print or retina scanning for access privileges.
- Intellectual Property: Trademark image registration, where a new candidate mark is compared with existing marks to ensure no risk of confusing property ownership.
- Medical Diagnosis: Using CBIR in a medical database of medical images to aid diagnosis by identifying similar past cases.
- Crime prevention: Automatic face recognition systems, used by police forces.

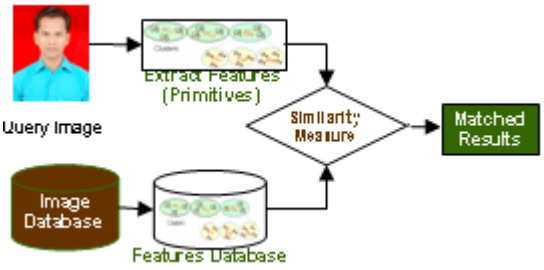


Fig. 1 CBIR (Content Based Image retrieval):Block Diagram.

Problem Statement

The problem involves entering an image as a query into a software application that is designed to employ CBIR techniques in extracting visual properties, and matching them. This is done to retrieve images in the database that are visually similar to the query image.

II. Feature Extraction Techniques.

2.1. Colour

2.1.1 Definition

One of the most important features that make possible the recognition of images by humans is colour. Colour is a property that depends on the reflection of light to the eye and the processing of that information in the brain. We use colour everyday to tell the difference between objects, places, and the time of day . Usually colours are defined in three dimensional colour spaces.

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

These could either be RGB (Red, Green, and Blue), HSV (Hue, Saturation, and Value) or HSB (Hue, Saturation, and Brightness). The last two are dependent on the human perception of hue, saturation, and brightness.

Most image formats such as JPEG, BMP, GIF, use the RGB colour space to store information. The RGB colour space is defined as a unit cube with red, green, and blue axes. Thus, a vector with three co-ordinates represents the colour in this space. When all three coordinates are set to zero the colour perceived is black. When all three coordinates are set to 1 the colour perceived is white. The other colour spaces operate in a similar fashion but with a different perception.

2.1.2 Methods of Representation

The main method of representing colour information of images in CBIR systems is through colour histograms. A colour histogram is a type of bar graph, where each bar represents a particular colour of the colour space being used. In MatLab for example you can get a colour histogram of an image in the RGB or HSV colour space. The bars in a colour histogram are referred to as bins and they represent the x-axis. The number of bins depends on the number of colours there are in an image. The y-axis denotes the number of pixels there are in each bin. In other words how many pixels in an image are of a particular colour.

An example of a colour histogram in the HSV colour space can be seen with the image shown. To view a histogram numerically one has to look at the colour map or the numeric representation of each bin. As one can see from the colour map each row represents the colour of a bin. The row is composed of the three coordinates of the colour space. The first coordinate represents hue, the second saturation, and the third, value, thereby giving HSV. The percentages of each of these coordinates are what make up the colour of a bin. Also one can see the corresponding pixel numbers for each bin, which are denoted by the blue lines in the histogram.

Quantization in terms of colour histograms refers to the process of reducing the number of bins by taking colours that are very similar to each other and putting them in the same bin. By default the maximum number of bins one can obtain using the histogram function in MatLab is 256. For the purpose of saving time when trying to compare colour histograms, one can quantize the number of bins. Obviously quantization reduces the information regarding the content of images but as was mentioned this is the trade-off when one wants to reduce processing time.

There are two types of colour histograms, Global colour histograms (GCHs) and Local colour histograms (LCHs). A GCH represents one whole image with a single colour histogram. An LCH divides an image into fixed blocks and takes the colour histogram of each of those blocks. LCHs contain more information about an image but are computationally expensive when comparing images. "The GCH is the traditional method for colour based image retrieval. However, it does not include information concerning the colour distribution of the regions" of an image. Thus when comparing GCHs one might not always get a proper result in terms of similarity of images.

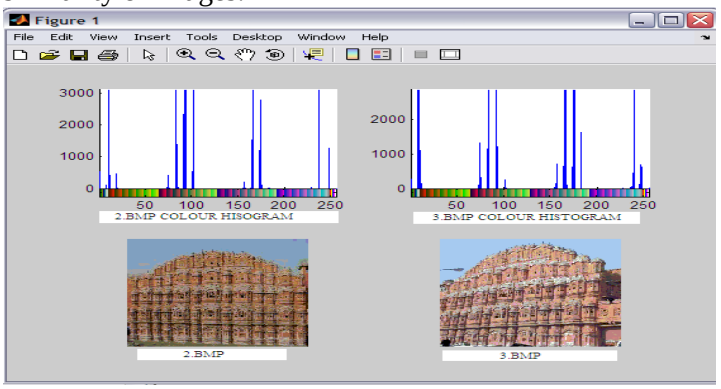


Fig. 2: Histogram of two Image of same class.

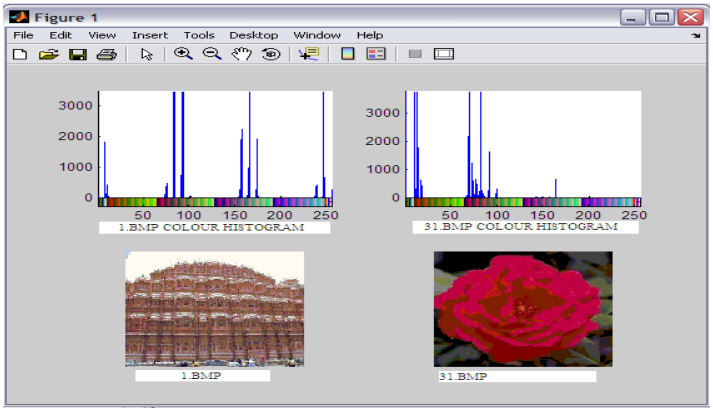


Fig. 3: Histogram of two Images of two different classes.

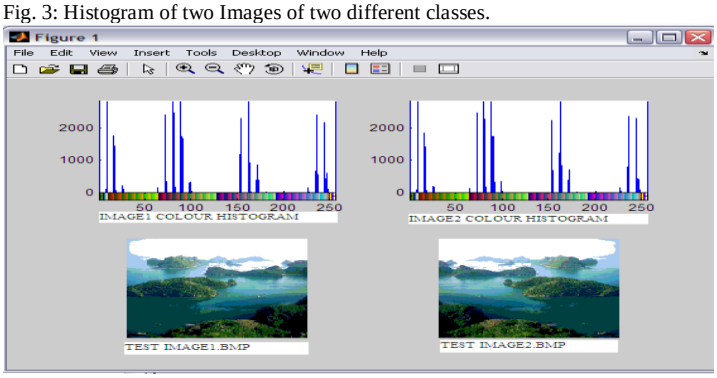


Fig. 4: Histogram of two Image of same class.

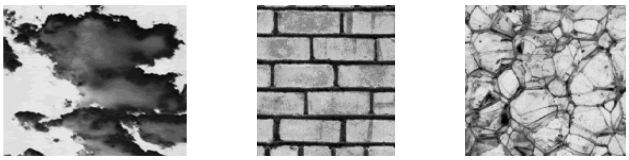
http://sites.google.com/site/ijcsis/
ISSN 1947-5500

2. 2 TEXTURE

Texture is an intuitive concept that describes properties like smoothness, coarseness, and regularity of a region. Texture is an important element to human vision, it provides cues to scene depth and surface orientation. In the next sections, Intensity-based texture will be described which has been the topic of investigation for many years and has proven useful. For example, the black and white television proves the usability of Intensity-based texture: people are able to see 3D in a 2D black and white screen. So, it seems important to look at Intensity-based textures before looking at colourful textures because the techniques used by Intensity-based textures can probably be expanded to colour-texture.

2.2.1 Texture Definition

Texture is that innate property of all surfaces that describes visual patterns, each having properties of homogeneity. It contains important information about the structural arrangement of the surface, such as; clouds, leaves, bricks, fabric, etc. It also describes the relationship of the surface to the surrounding environment. In short, it is a feature that describes the distinctive physical composition of a surface.
Texture may be defined as a local arrangement of image irradiances projected from a surface patch of perceptually homogeneous irradiances. Texture regions give different interpretations at different distances and at different degrees of visual attention. At a standard distance with normal attention, it gives the notion of macro-regularity that is characteristic of the particular texture. When viewed closely and attentively, homogeneous regions and edges, sometimes constituting texels, are noticeable.



(a) Clouds (b) Bricks (c) Rocks

Figure 5: Examples of Texture

2.2.3Texture properties include:

- Coarseness
- Contrast
- Directionality

Moment	Expression.	Description
Mean	$m = \sum_{i=0}^{L-1} Z_i P(Z_i)$	To measure the average intensity
Standard deviation	$\sigma = \sqrt{\mu_2(Z)} = \sqrt{\sigma^2}$	To measure the average contrast
Smoothness	$R = 1 - \frac{1}{(1+\sigma^2)}$	To measure the relative smoothness of the intensity in a region.
Third moment	$\mu_3 = \sum_{i-0}^{L-1} (Z_i - m)^3 p(Z_i)$	To measure the skewness of a histogram
Uniformity	$U = \sum_{i=0}^{L-1} P^2(Z_i)$	To measure the uniformity
Entropy	$e = -\sum_{i=0}^{L-1} p(Z_i) \log_2 P(Z_i)$	To measure the randomness

Table: 1 statistical parameters

- Line-likeness
- Regularity
- Roughness

Texture is one of the most important defining features of an image. It is characterized by the spatial distribution of grey levels in a neighbourhood. In order to capture the spatial dependence of gray-level values, which contribute to the perception of texture, a two-dimensional dependence texture analysis matrix is taken into consideration.
This two-dimensional matrix is obtained by decoding the image file; jpeg, bmp, etc.

2.2.4 Texture Features:

Methods of Representation

There are three principal approaches used to describe texture; statistical, structural and spectral.

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

- Statistical techniques characterize textures using the statistical properties of the grey levels of the points/pixels comprising a surface image. Typically, these properties are computed using: the grey level co-occurrence matrix of the surface, or the wavelet transformation of the surface.
- Structural techniques characterize textures as being composed of simple primitive structures called "texels" (or texture elements). These are arranged regularly on a surface according to some surface arrangement rules.
- Spectral techniques are based on properties of the Fourier spectrum and describe global .periodicity of the grey levels of a surface by Identifying high-energy peaks in the Fourier spectrum.

For optimum classification purposes, what concern are the statistical techniques of characterization. This is because it is these techniques that result in computing texture properties. The most popular statistical representations of texture are:

- Co-occurrence Matrix
- Tamura Texture
- Wavelet Transform.

2.3. SHAPE

2.3. 1 Definition of Shape:

Shape may be defined as the characteristic surface configuration of an object; an outline or contour. It permits an object to be distinguished from its surroundings by its outline. Shape representations can be generally divided into two categories:

Boundary-based, and Region-based.

Figure 6: Boundary-based & Region-based

Boundary-based shape representation only uses the outer boundary of the shape. This is done by describing the considered region using its external characteristics; i.e., the pixels along the object boundary. Region-based shape representation uses the entire shape region by describing the considered region using its internal characteristics; i.e., the pixels contained in that region.

2.3.2 Methods of Representation

For representing shape features mathematically, we have:

Boundary-based:
Polygonal Models, boundary partitioning
Fourier Descriptors:
Splines, higher order constructs
Curvature Models
Region-based:
Super quadrics
Fourier Descriptors
Implicit Polynomials
Blum's skeletons

The most successful representations for shape categories are Fourier Descriptor and Moment Invariants: The main idea of Fourier Descriptor is to use the Fourier transformed boundary as the shape feature.
The main idea of Moment invariants is to use region-based moments, which are invariant to transformations as the shape feature.

III. IMAGE FEATURE MATCHING

3.1 Similarity Distance measures

CBIR employs low level image features such as color, shape or texture to achieve objective and automatic indexing, in contrast to subjective and manual indexing in traditional image indexing. For contend based image retrieval, the image feature extracted is usually an N-dimensional feature vector which can be regarded as a point in a N-dimensional space. Once images are indexed into the database using the extracted feature vectors, the retrieval of images is essentially the determination of similarity between the features of query image and the features of target images in database, which is essentially the determination of distance between the feature vectors representing the images. The desirable distance measure should reflect human perception. That is to say, perceptually similar images should have smaller distance between them and perceptually different images should have larger distance between them. Therefore, for a given shape feature, the higher the retrieval accuracy, the better the distance measure. Various distance measures have been exploited in image retrieval, are discussed below.

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

3.2 SIMILARITY MEASUREMENTS

A similarity measurement is normally defined as a metric distance. In this section different similarity measurements are described in details.

Minkowski-form distance metrics

The Minkowski metric between two point's p = (x1, y1) and q = (x2, y2) is defined as:

$$d^k(P,Q) = (|x_1 - y_1|^K + |x_2 - y_2|^K \quad \textbf{(1)}$$

The histogram distance

The histogram distance, calculated per bin m, between a query image q and a target image t is denoted as:

$$D_i(q,t) = \sum_{m=0}^{M-1} |h_q[m] - h_t[m]| \quad \textbf{(2)}$$

Where M is the total number of bins, hq is the normalized query histogram, and ht is the normalized target histogram. We recognize Di(q; t) as the Minkowski form metric with k=1.

Histogram the Euclidean distance

The Euclidean distance is a Minkowski form with k=2:

$$D_e(q,t) = \sqrt{\sum_{m=0}^{M-1} (h_q[m] - h_t[m])^2} \quad \textbf{(3)}$$

The distances (i.e., calculated Minkowski-form distance measures) only take account for the correspondence between each histogram bin. and do not make use of information across bins. This issue has been recognized in histogram matching. As a result, quadratic distance is proposed to take similarity across dimensions into account. It has been reported to provide more desirable result than only matching between similar bins of the color histograms. However, since the histogram quadratic distance computes the cross similarity between colours, it is computationally expensive.

Histogram quadratic distance

The quadratic-form distance between two feature vectors q and t is given by:

$$D_e(q,t) = (h_q - h_t)^T A(h_q - h_i) \quad \textbf{(4)}$$

Where A = [aij] is a similarity matrix. aij denotes the similarity between elements with indexes i and j. Please note, that h_q and ht are denoted as vectors.

In order to determine the intersection similarity (S) we adapt Equation to give:

$$s_{q,t} = \sum_{m=0}^{M-1} 1 - |h_q(m) - h_t(m)|. \quad \textbf{(5)}$$

Minkowski-form distance metrics compare only the same bins between color histograms and are defined as:

$$d(Q,I) = \sum_{i=1}^{N} |H_Q[i] - H_I[i]|^r \quad \textbf{(6)}$$

Where Q and I are two images ,N , is the number of bins in the color histogram (for each image we reduce the colours to N , in the RGB color space, so each color histogram has N bins, HQ[i] is the value of bin i in color histogram HQ , which represents the image Q , and HI[i] is the value of bin i in color histogram HI , which represents the image I.

When r=1, the Minkoski-form distance metric becomes L1. When r=2 , the Minkoski-form distance metric becomes the Euclidean distance. In fact, this Euclidean distance can be treated as the spatial distance in a multi-dimensional space.

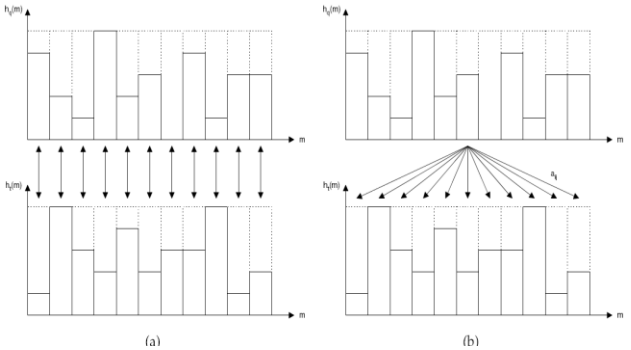


Figure 7 : (a) Minkowski-form distance metrics compare only similar bins between histograms.
(b) Quadratic-form distance metrics compare multiple bins between histograms using similarity matrix A = [aij].

Quadratic-form distance metrics

The quadratic distance, also called cross distance, is used in the QBIC-system. This method considers the cross-correlation between histogram bins based on the perceptual similarity of the colours represented by the bins. The set of correlation values is represented in a similarity matrix.

Quadratic-form distance metric compares not only the same bins but multiple bins between color histograms and is defined as:

$$d(Q,I) = (H_Q - H_I)^t A(H_Q - H_I) \quad \textbf{(7)}$$

Where Q and I are two images HQ is the color histogram of image Q, HI is the color histogram of image I, |A|=[ai,j] is a NXN 54 , matrix, N is the number of bins in the color histograms, and ai,j denotes the similarity between colours i and j.The similarity matrix is obtained through a complex algorithm:

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

$$\alpha_{q,i} = 1 - \frac{[(v_q - v_i)^2 + (s_q \cos(h_q) - s_i \cos(h_i))^2 + (s_q \sin(h_q) - s_i \sin(h_i))^2]^{\frac{1}{2}}}{\sqrt{5}}$$

where (hq, sq, vq) and (hi si , vi) represent hue, saturation, and value components for two colours indexed by two histogram bins. Quadratic-form distance metrics overcome a shortcoming of the Minkoski-form distance metrics in that the latter assumes that bins in color histograms are totally unrelated, while the former does not.

Chebyshev Distance

This distance calculation metric is named after Pafnuty Lvovich Chebyshev, it is also known as chessboard distance, the equation is defined as

$$d = \max(|x2 - x1|, |y2 - y1|) \quad \textbf{(8)}$$

Bray Curtis distance

$$d_{i,j} = \frac{\sum_{k=1}^{m} |x_{ik} - x_{jk}|}{\sum_{k=1}^{m} |x_{ik} + x_{jk}|} \quad \textbf{(9)}$$

Manhattan Distance

Manhattan distance is also known as Taxicab distance. This is because it comes from the fact that it represents the shortest distance a car will drive in a city laid out in square blocks. For example, in the plane, the Manhattan distance between the point P1 with coordinates (x1, y1) and the point P2 at (x2, y2) is

$$|x_1 - x_2| + |y_1 - y_2| \quad \textbf{(10)}$$

Hamming Distance

$$HD = \frac{1}{N} \sum_{j=1}^{N} X_j (XOR) Y_j \quad \textbf{(11)}$$

IV SYSTEM PERFORMANCE EVALUATION.

Precision and Recall:

Testing the effectiveness of the image search engine is about testing how well can the search engine retrieve similar images to the query image and how well the system prevents the returned results that are not relevant to the source at all in the user point of view. The big question here is how we know that which image is relevant. Determining whether or not two images are similar is purely up to the user's perception. Human perceptions can easily recognise the similarity between two images although in some cases, different users can give different opinions. Two evaluation measures were used here to evaluate the effectiveness of the image search engine system.

The first measure is Recall. It is a measure of the ability of a system to present all relevant items. The equation for calculating recall is given below:

$$\text{Recall} = \frac{\text{number of relevant items retrieved}}{\text{number of relevant items in collection}} \quad \textbf{(12)}$$

The second measure is Precision. It is a measure of the ability of a system to present only relevant items. The equation for calculating precision is given below.

$$\text{Precision} = \frac{\text{number of relevant items retrieved}}{\text{total number of items retrieved}} \quad \textbf{(13)}$$

A retrieval score

A retrieval score can be computed according to the following evaluation criterion: for each query, the system returns the 'x' closest images to the query, including the query image itself (as the distance from the query image to itself is zero). The number of mismatches can be computed as the number of images returned that belong to a class different than that of the query image, in addition to the number of images that belong to the query image class, but that have not been returned by the system. The retrieval score for one class can be then computed as

$$\text{Re}trivalScore = 100 \times \left[1 - \left(\frac{mismatches}{x}\right)\right]\% \quad \textbf{(14)}$$

V Conclusions

CBIR at present is still topic of research interest. Different features are used for retrieval of images such as Image colour quadratic distance for image histogram, Image Euclidian distance for image wavelet transform; image Hamming Distance. And corresponding retrieval Recall and Precision parameters are calculated for each feature. For as to increase retrieval efficiency combination of these features should be used instead of using a single feature for image retrieval.

The retrieval efficiency and timing performance can be further increased if the image collection is trained

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

(pre-processed) and grouped using supervised learning such as classification or unsupervised learning such as clustering. With that, the image with high similarities in the feature space will be group together and result a smaller search space. This will greatly enhance the search time and precision.

Future work

Classification and Clustering:

The retrieval efficiency and timing performance can be further increased if the image collection is trained (pre-processed) and grouped using supervised learning such as classification or unsupervised learning such as clustering. With that, the image with high similarities in the feature space will be group together and result a smaller search space. This will greatly enhance the search time and precision.

Acknowledgements

I would like to take this opportunity to thank my Guide, Prof. S. N. Kore ,Head of electronics Engineering Department Walchand College of Engineering, Sangli for his valuable guidance, constant inspiration, advice, and encouragement throughout the course of this work. I am also thankful to him for his timely and helpful advices to understand the nature of topic and careful reviews on the subject.

Also, I would like to our Principal Dr. S. A. Halkude and Head Dr. S. K. Dixit for their continuous encouragement.

References

1. Lenina Birgale, Manesh Kokare and Dharmpal Doye, “Colour and Texture Features for Content Based Image Retrieval”, Proceedings of the International Conference on Computer Graphics, Imaging and Visualisation (CGIV'06) 0-7695-2606-3/06 $20.00 © 2006IEEE.
2. ZHEN-HUA ZHANG, YONG QUAN, WEN-HUI LI, WU GUO, “A NEW CONTENT-BASED IMAGE RETRIEVAL”, Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian, 13-16 August 2006.
3. P.S.Hiremath, S. Shivashankar, and Jagadeesh Pujari, “WAVELET BASED FEATURES FOR COLOR TEXTURE CLASSIFICATION WITH APPLICATION TO CBIR”, IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.9A, September 2006.
4. P.S.Hiremath, S.Shivashankar, ” WAVELET BASED FEATURES FOR TEXTURE CLASSIFICATION”, GVIP Journal, Volume 6, Issue 3, December, 2006
5. Subrahmanyam Murala, Anil Balaji Gonde, R. P. Maheshwari,” Color and Texture Features for Image Indexing and Retrieval”, 2009 IEEE International Advance Computing Conference (IACC 2009) Patiala, India, 6-7 March 2009.
6. L. Kotoulas and I. Andreadis, “Colour histogram content-based image retrieval and hardware implementation”, IEE Proc.-Circuits Devices Syst., Vol. 150, No. 5, October 2003.
7. T F. Ang , 2 Y K. Cheong, 3 L Y. Por, 4 K K. Phang, “Colour-Based Image Retrieval Using Global Colour Histogram and Harbin”, Multimedia Cyberscape Journal, Volume 5, Number 1, Year 2007.
8. Igor Marinovi and Igor Fürstner Manufaktura d.o.o., Subotica, Serbia “Content-based Image Retrieval ”, 1-4244-2407-8/08/$20.00 ©2008 IEEE.
9. Dengsheng Zhang and Guojun Lu , “EVALUATION OF SIMILARITY MEASUREMENT FOR IMAGE RETRIEVAL”, 1-4240-2407-8/08/$25.00 ©2004 IEEE
10. Greg Pass Ramin Zabih*, “Histogram Refinement for Content-Based Image Retrieval”, 0-8186-7620-5/96 $5.00 0 1996 IEEE.
11. Gonzalez, R.C., Woods, R.E, “Digital Image Processing” 2nd Ed., Prentice Hall
12. Robi Polikar, “The Engineers Ultimate Guide To Wavelet Analysis”, [Online Document], Available: http://engineering.rowan.edu/~polikar/WAVELETS/WTtutorial.html
13. VisTex, “color image database”, at http://www.white.media.mit.edu/vismod/imagery/ vision
14. Texture, MIT media Lab.
15. MATLAB Manual

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

Mr. Kondekar Vipul H. received Master of Engineering in Electronics Engineering with specialization in Computer engineering from Walchand College of Engineering, Sangali Maharastra-India He is working as Lecturer for Last 6 Years in Electronics and Telecommunication Engineering Department at Walchand institute of Technology, Solapur University, Solapur, maharastra-India. His current research interest areas are Image Processing and Microcontroller based system design. He has authored and co-authored more than 10 technical papers published in various prestigious national/international journals and referred conference,symposium,workshop proceedings.

Mr. Kolkure Vijaykumar S. has completed Master of Engineering (ME) in Electronics Engineering with specialization in Computer engineering from Walchand College of Engineering; Sangli Maharastra-India .He has 3 years' experience of working in IT industry. Presently he is working as Lecturer for Last 1 Years in Electronics and Telecommunication Engineering Department at Bharat Ratna Indira Gandhi College of Engineering, Solapur University, Solapur, Maharastra-India. His current research interest areas are Image Processing and Embedded systems. He has more than 8 technical papers published in various prestigious national/international journals and referred conference/symposium/workshop proceedings.

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

AHB Compatible DDR SDRAM Controller IP Core for ARM BASED SOC

[1]Dr.R.Shashikumar, [2]C.N. Vijay Kumar, [3]M.Nagendrakumar, [4]C.S.Hemanthkumar
Professor Asst.prof Asst.Prof Sr.Lecturer
ECE dept, SJCIT, Chikkaballapur, Karnataka, India JVIT, Bidadi
,

***Abstract*— DDR SDRAM is similar in function to the regular SDRAM but doubles the bandwidth of the memory by transferring data on both edges of the clock cycles.DDR SDRAM most commonly used in various embedded application like networking, image/video processing, Laptops ete. Now a day's many applications needs more and more cheap and fast memory. Especially in the field of signal processing, requires significant amount of memory. The most used type of dynamic memory for that purpose is DDR SDRAM. For FPGA design the IC manufacturers are providing commercial memory controller IP cores working only on their products. Main disadvantage is the lack of memory access optimization for random memory access patterns. The 'data path' part of those controllers can be used free of charge. This work propose an architecture of a DDR SDRAM controller, which takes advantage of those available and well tested data paths and can be used for any FPGA device or ASIC design.[5]. In most of the SOC design, DDR SDRAM is commonly used. ARM processor is widely used in SOC's; so that we focused to implement AHB compatible DDR SDRAM controller suitable for ARM based SOC design.**

Keywords-AHB; DDR SDRAM; Verilog;IP core;

1 INTRODUCTION

The DDR SDRAM is a high-speed CMOS, dynamic random-access memory. It is internally configured as a quad bank DRAM. The DDR SDRAM uses double data rate architecture to achieve high-speed operation. The double data rate architecture is essentially 2nprefetch architecture with an interface designed to transfer two data words per clock cycle at the I/O pins. A single read or write access for the DDR SDRAM effectively consists of a single 2*n*-bit wide, one-clock-cycle data transfer at the internal DRAM core and two corresponding *n*-bit wide, one-half clock-cycle data transfers at the I/O pins. A bidirectional data strobe (DQS) is transmitted externally, along with data, for use in data capture at the receiver. DQS is a strobe transmitted by the DDR SDRAM during READs and by the memory controller during WRITEs. DQS is edge-aligned with data for READs and center-aligned with data for WRITEs.

Read and write accesses to the DDR SDRAM are burst oriented; accesses start at a selected location and continue for a programmed number of locations in a programmed sequence. Accesses begin with the registration of an ACTIVE command, which is then followed by a READ or WRITE command. The address bits registered coincident with the ACTIVE command are used to select the bank and row to be accessed. The address bits registered coincident with the READ or WRITE command are used to select the bank and the starting column location for the burst access.

The DDR SDRAM provides for programmable READ or WRITE burst lengths of 2, 4, or 8 locations. An auto precharge function may be enabled to provide a self timed row precharge that is initiated at the end of the burst access This model has implemented in RTL by Verilog. The focus of this work is to implement behavioral model of DDR SDRAM and also implemented on the Xilinx Spartan series FPGA.The Top level model is as shown in Fig.1. The core contains mainly two parts, AHB Slave and DDR SDRAM controller.

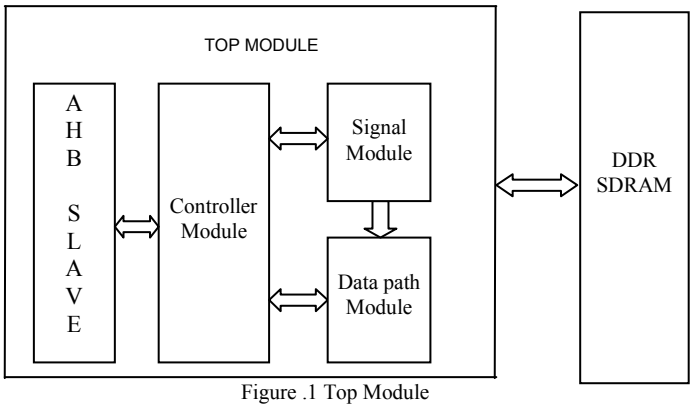


Figure .1 Top Module

2. FUNCTIONAL BLOCK DIAGRAM

The functional block diagram of the DDR controller is shown in Fig.2. It consists of four modules, AHB Salve, the main control module, the signal generation module and the data path module. The AHB slave normally connected to the AHB bus Arbiter in the design. AHB master sends data to the AHB slave based on that protocol. Burst mode read and writes and split transaction read and writes transactions supported. AHB slave complies with the processor interface protocol of ARM processor. The main control module has two state machines and a refresh counter, which generates proper istate and cstate outputs according to the system interface control signals. The signal generation module generates the address and command signals required for DDR based on istate and cstate. The data path module performs the data latching and dispatching of the data between the Processor and DDR.

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

The DDR SDRAM provides for programmable READ or WRITE burst lengths of 2, 4, or 8 locations. An AUTO PRECHARGE function may be enabled to provide a self-timed row precharge that is initiated at the end of the burst access. As with standard SDR SDRAMs, the pipelined, multibank architecture of DDR SDRAMs allows for concurrent operation, thereby providing high effective bandwidth by hiding row precharge and activation time. An auto refresh mode is provided, along with a power-saving power-down mode.

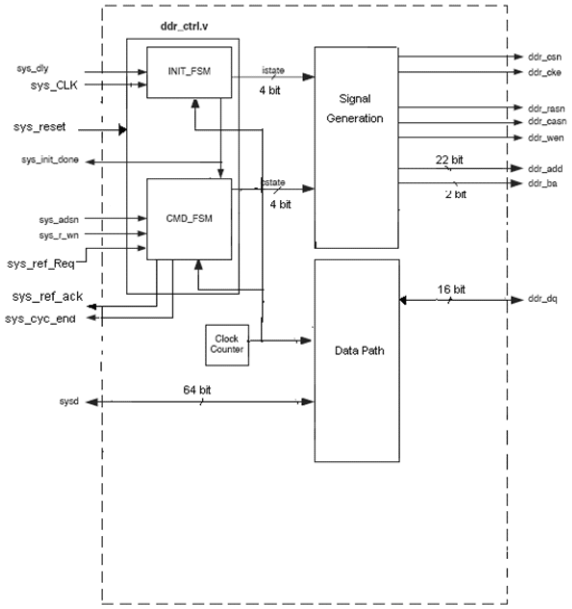


Figure .2 Functional Block Diagram

3. FUNCTIONAL DESCRIPTION

3.1. Initialization

DDR SDRAMs must be powered up and initialized in a predefined manner. Operational procedures other than those specified may result in undefined operation.
DDR SDRAM requires a 200µs delay prior to applying an executable command [7]. Once the 200µs delay has been satisfied, a DESELECT or NOP command should be applied. Following the NOP command, a PRECHARGE ALL command should be applied. Next a LOAD MODE REGISTER command should be issued. A PRECHARGE ALL command should then be applied, placing the device in the all banks idle state. Once in the idle state, two AUTO REFRESH cycles must be performed (tRFC must be satisfied.) Additionally, a LOAD MODE REGISTER command for the mode register is issued [16].

3.2 Register Definition

The mode register is used to define the specific mode of operation of the DDR SDRAM. This definition includes the selection of a burst length, a burst type, a CAS latency and an operating mode. The mode register is programmed via the MODE REGISTER SET command (with BA0 = 0 and BA1 = 0) and will retain the stored information until it is programmed again or the device loses power (except for bit A8, which is self-clearing). Reprogramming the mode register will not alter the contents of the memory, provided it is performed correctly. The mode register must be loaded (reloaded) when all banks are idle and no bursts are in progress, and the controller must wait the specified time before initiating the subsequent operation. Violating either of these requirements will result in unspecified operation. Mode register bits A0-A2 specify the burst length, A3 specifies the type of burst (sequential or interleaved), A4-A6 specifies the CAS latency, and A7-A12 specifies the operating mode. Fig.3 shows the description of mode register.

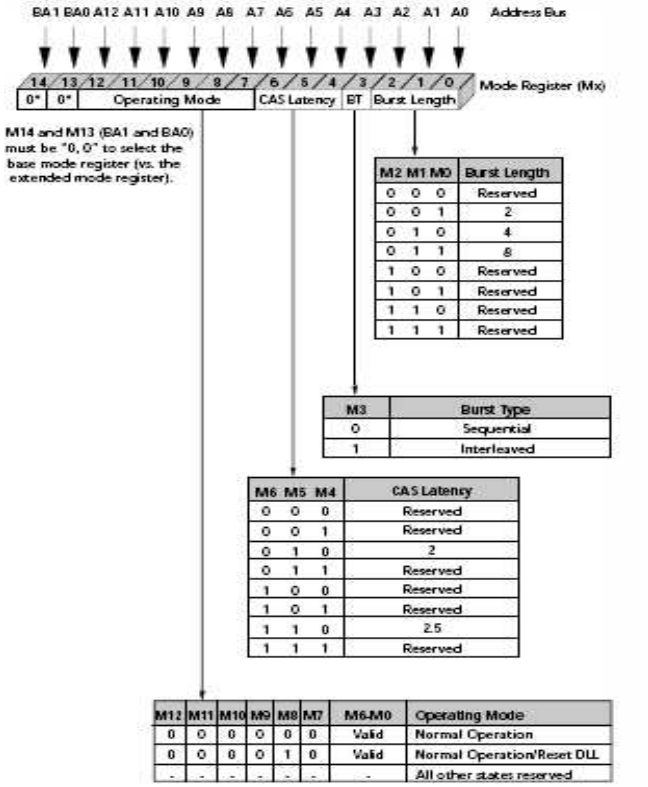


Figure.3 Mode register definition

3.3 DDR Commands

Table 1 presents the commands issued by the controller. These commands are passed to the memory using the following control signals[7]:
• Row Address Select (RAS)
• Column Address Select (CAS)
• Write Enable (WE)
• Clock Enable (CKE) (always held High after device configuration)
• Chip Select (CS) (always held Low during device operation)

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

Signal No.	Function	$\overline{RAS}$	$\overline{CAS}$	$\overline{WE}$
1	Load Mode Register	L	L	L
2	Auto Refresh	L	L	H
3	Precharge(1)	L	H	L
4	Select Bank Activate Row	L	H	H
5	Write Command	H	L	L
6	Read Command	H	L	H
7	No Operation (NOP)	H	H	H

Table 1 DDR SDRAM Commands

3.3.1 Command Functions

The fallowing commands are used in DDR SDRAM controller core.

- Mode Register **:**

The Mode register is used to define the specific mode of DDR SDRAM operation, including the selection of burst length, burst type, CAS latency, and operating mode.

- Auto Refresh **:**

The REFRESH command instructs the controller to perform an AUTO REFRESH command to the SDRAM. The controller will acknowledge the REFRESH command with ACK. DDR SDRAM is somewhat similar to regular SDRAM. Both will break the RAM into smaller chunks for simultaneous, synchronized request-and-reply access. In addition, both types of memory can be packaged in DIMM modules. However, DDR SDRAM will perform the alternating request-and-reply rhythm on both the rise and fall of the clock cycle. This method effectively doubles the bandwidth available and increases the speed the system can access data in memory

- Precharge:

The PRECHARGE command is used to deactivate the open row in a particular bank. The bank is available for subsequent row activation for a specified time (tRP) after the PRECHARGE command is issued. Input A10 determines whether one or all banks are precharged.

- ACTIVE Command:

The ACTIVE command activates a row in a bank, allowing any READ or WRITE commands to be issued to a bank in the memory array. After a row has been opened, READ or WRITE commands can be issued to that row, subject to the tRCD specification. When the controller detects an incoming address that refers to a row in a bank other than the currently opened row, the controller issues an address conflict signal. A PRECHARGE command is also issued by the controller to deactivate the open row. The controller also issues another ACTIVE command to the new row.

- READ Command**:**

The READ command is used to initiate a burst read access to an active row. The value on BA0 and BA1 selects the bank address. The address inputs provided on A0 – Ai select the starting column location. After the read burst is over, the row is still available for subsequent access until it is precharged.

- WRITE Command**:**

The WRITE command is used to initiate a burst access to an active row. The value on BA0 and BA1 selects the bank address, while the value on address inputs A0 – Ai selects the starting column location in the active row. The value of Write Latency is equal to one clock cycle.

4 .DDR SDRAM Controller Block

The controller block mainly consists of two FSMs

- Initial FSM
- Command FSM

It generates the required commands for initializing the DDR SDRAM. The block diagram for the DDR SDRAM controller is as shown in Fig.4.

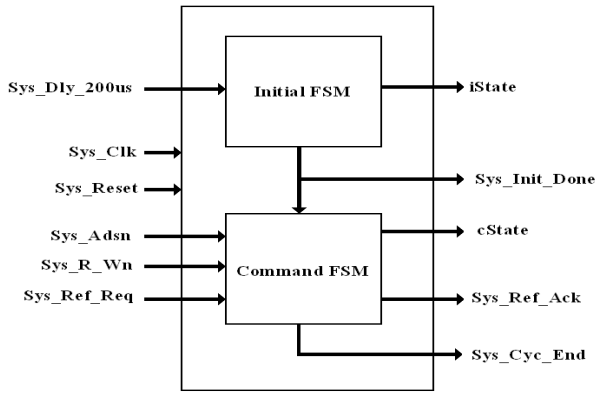


Figure 4 Block Diagram of Controller Block

4.1 DDR SDRAM Initial FSM

Before normal memory accesses can be performed, DDR needs to be initialized by a sequence of commands. The INIT_FSM state machine handles this initialization. Fig.5 shows the state diagram of the INIT_FSM state machine. During reset, the INIT_FSM is forced to the i_IDLE state. After reset, the sys_dly_200US signal will be sampled to determine if the 200µs power/clock stabilization delay is completed. After the power/clock stabilization is complete, the DDR initialization sequence will begin and the INIT_FSM will switch from i_IDLE to i_NOP state and in the next clock to I_PRE.

The initialization starts with the PRECHARGE ALL command. Next a LOAD MODE REGISTER command will be applied for the extended mode register to enable the DLL inside DDR, followed by another LOAD MODE REGISTER command to the mode register to reset the DLL. Then a PRECHAGE command will be applied to make all banks in the device to idle state. Then two, AUTO REFRESH commands, and then the LOAD MODE REGISTER command to configure DDR to a specific mode of operation. After issuing the LOAD MODE REGISTER command and the tMRD timing delay is satisfied, INIT_FSM goes to i_ready state and remains there for the normal memory

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

access cycles unless reset is asserted. Also, signal sys_init_done is set to high to indicate the DDR initialization is completed. The i_PRE, i_AR1, i_AR2, i_EMRS and i_MRS states are used for issuing DDR commands. The LOAD MODE REGISTER command configures the DDR by loading data into the mode register through the address bus. The data present on the address bus (ddr_add) during the LOAD MODE REGISTER command is loaded to the mode register. The mode register contents specify the burst length, burst type, CAS latency, etc. A PRECHARGE/AUTO PRECHARGE command moves all banks to idle state. As long as all banks of the DDR are in idle state, mode register can be reloaded with different value thereby changing the mode of operation. However, in most applications the mode register value will not be changed after initialization. This design assumes the mode register stays the same after initialization.

4.1.1 Initial FSM State Diagram:

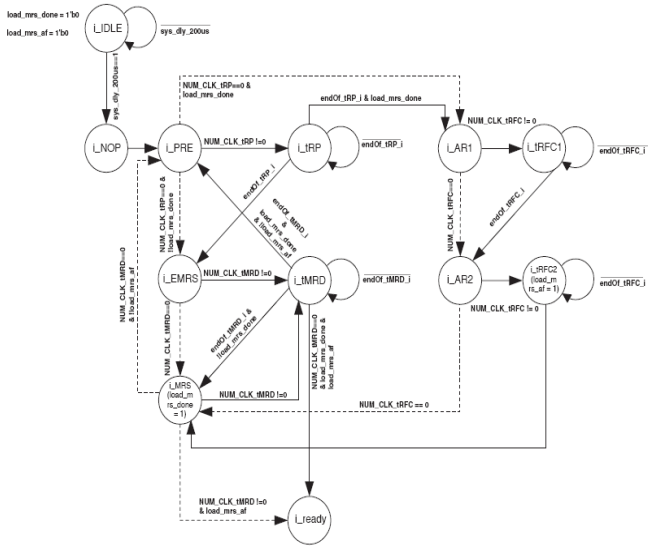


Figure .5 Initial FSM State Diagram

As mentioned above, certain timing delays (tRP, tRFC, tMRD) need to be satisfied before another non-NOP command can be issued. These SDRAM delays vary from speed grade to speed grade and sometimes from vendor to vendor. To accommodate this without sacrificing performance, the designer needs to modify the HDL code for the specific delays and clock period (tCK). According to these timing values, the number of clocks the state machine will stay at i_tRP, i_tRFC1, i_tRFC2, i_tMRD states will be determined after the code is synthesized. In cases where tCK is larger than the timing delay, the state machine doesn't need to switch to the timing delay states and can go directly to the command states[15]. The dashed lines in Fig 5 show the possible state switching paths.

4.1.2 Different states of Initial FSM

a) Idle:

When reset is applied the initial fsm is forced to IDLE state irrespective of which state it is actually in when system is in idle it remains idle without performing any operations.

b) No Operation:

The NO OPERATION (NOP) command is used to instruct the selected DDR SDRAM to perform a NOP (CS# is LOW with RAS#, CAS#, and WE# are HIGH). This prevents unwanted commands from being registered during idle or wait states. Operations already in progress are not affected.

c) Precharge:

The PRECHARGE command is used to deactivate the open row in a particular bank or the open row in all banks as shown in Fig.6. The value on the BA0, BA1 inputs selects the bank, and the A10 input selects whether a single bank is precharged or whether all banks are precharged.

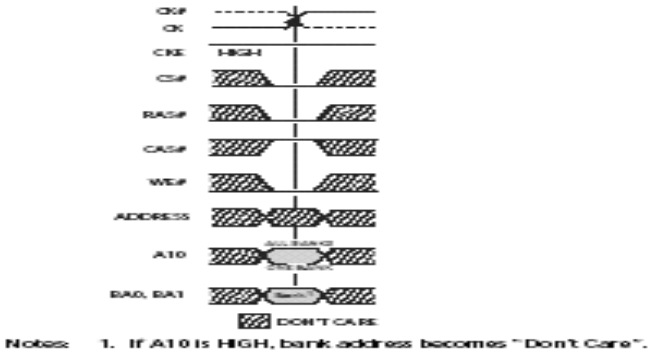


Figure .6 Precharge Commands.

d) Auto Refresh:

AUTO REFRESH is used during normal operation of the DDR SDRAM and is analogous to CAS#-before-RAS# (CBR) refresh in DRAMs. This command is nonpersistent, so it must be issued each time a refresh is required. All banks must be idle before an AUTO REFRESH command is issued.

e) Load Mode register(LMR):

The mode registers are loaded via inputs A0–A*n*. The LOAD MODE REGISTER command can only be issued when all banks are idle, and a subsequent executable command cannot be issued until tMRD is met.

f) Read/Write Cycle:

The Fig.5 shows the state diagram of CMD_FSM which handles the read, write and refresh of the SDRAM.

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

The CMD_FSM state machine is initialized to c_idle during reset. After reset, CMD_FSM stays in c_idle as long as sys_INIT_DONE is low which indicates the SDRAM initialization sequence is not yet completed. Once the initialization is done, sys_ADSn and sys_REF_REQ will be sampled at the rising edge of every clock cycle. A logic high sampled on sys_REF_REQ will start a SDRAM refresh cycle. This is described in the following section. If logic low is sampled on both sys_REF_REQ and sys_ADSn, a system read cycle or system write cycle will begin. These system cycles are made up of a sequence of SDRAM commands. Initialization:

Prior to normal operation, DDR SDRAMs must be powered up and initialized in a predefined manner. Operational procedures, other than those specified, may result in undefined operation.

4.2 DDR SDRAM COMMAND FSM

The fig.7 shows the state diagram of CMD_FSM, which handles read, write and refresh of the DDR. The CMD_FSM state machine is initialized to c_idle during reset. After reset, CMD_FSM stays in c_idle as long as sys_init_done is low which indicates th e DDR initialization sequence is not yet completed. From this state, a READA/WRITEA/REFRESH cycle starts depending upon sys_adsn/ rd_wr_req_during_ ref_req signals as shown in the state diagram. All rows are in the "closed" status after the DDR initialization. The rows need to be "opened" before they can be accessed. However, only one row in the same bank can be opened at a time. Since there are four banks, there can be at most four rows opened at the same time. If a row in one bank is currently opened, it needs to be closed before another row in the same bank can be opened. ACTIVE command is used to open the rows and PRECHARGE is used to close the rows. When issuing the commands for opening or closing the rows, both row address and bank address need to be provided.

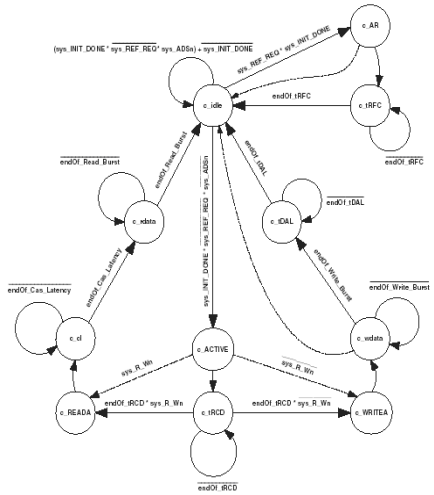


Figure .7 Command FSM State Diagram

In this design, the ACTIVE command will be issued for each read or write access to open the row. After a tRCD delay is satisfied, READA or WRITEA commands will be issued with a high ddr_add[10] to enable the AUTO REFRESH for closing the row after access. Therefore, the clocks required for read/write cycle are fixed and the access can be random over the full address range. Read or write is determined by the sys_r_wn status sampled at the rising edge of the clock before the tRCD delay is satisfied. If logic high is sampled, the state machine switches to c_READA. If a logic low is sampled, the state machine switches to c_WRITEA.

For read cycles, the state machine switches from, c_READA to c_cl for CAS latency, then switches to crate for transferring data from DDR to processor. The burst length determines the number of clocks the state machine stays in c_rdata state. After the data is transferred, it switches back to c_idle.

For write cycles, the state machine switches from c_WRITEA to c_wdata for transferring data from bus master to DDR, then switches to c_tDAL. After the clock rising edge of the last data in the burst sequence, no commands other than NOP can be issued to DDR before tDAL is satisfied.

4.2.1Different states of Command FSM

a) Refresh Cycle:

DDR memory needs a periodic refresh to hold the data. This periodic refresh is done using AUTO REFRESH command. All banks must be idle before an AUTO REFRESH command is issued. In this design all banks will be in idle state, as every read/write operation uses auto pre charge.

b) Active:

The ACTIVE command is used to open (or activate) a row in a particular bank for a subsequent access, like a read or a write, as shown in Fig.8. The value on the BA0, BA1 inputs selects the bank, and the address provided on inputs A0–A*n* selects the row.

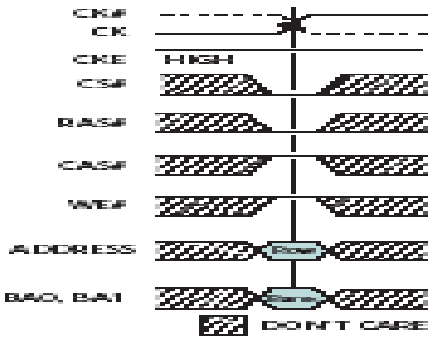


Figure.8 Activating a Specific Row in a Specific Bank.

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

c) Read:

The READ command is used to initiate a burst read access to an active row, as shown in Fig.8. The value on the BA0, BA1 inputs selects the bank, and the address provided on inputs A0–A*i* (where A*i* is the most significant column address bit for a given density and configuration) selects the starting column location.

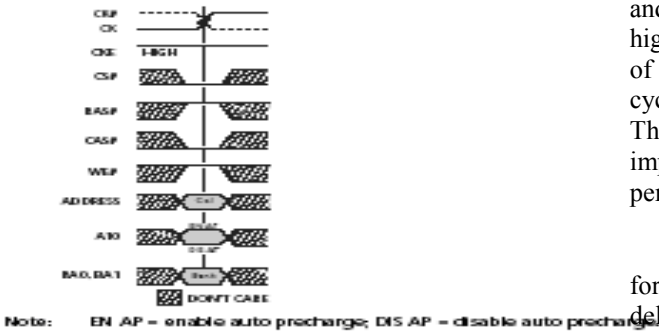


Figure .9 Read Command

d) Write:

The WRITE command is used to initiate a burst write access to an active row as shown in Fig.10. The value on the BA0, BA1 inputs selects the bank, and the address provided on inputs A0–A*i* (where A*i* is the most significant column address bit for a given density and configuration) selects the starting column location.

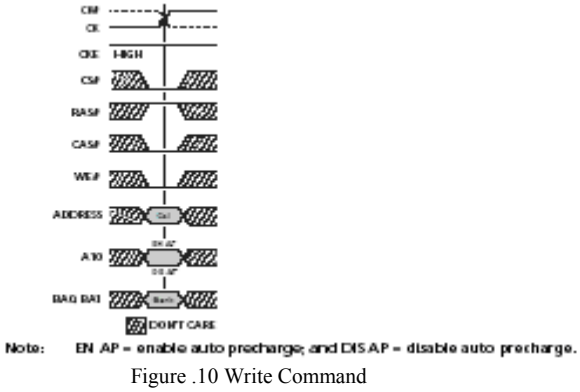


Figure .10 Write Command

Similar to the FP and EDO DRAM, row address and column address are required to pinpoint the memory cell location of the SDRAM access. Since SDRAM is composed of four banks, bank address needs to be provided as well.

The SDRAM can be considered as a four by N array of rows. All rows are in the "closed" status after the SDRAM initialization. The rows need to be "opened" before they can be accessed. However, only one row in the same bank can be opened at a time. Since there are four banks, there can be at most four rows opened at the same time. If a row in one bank is currently opened, it must be closed before another row in the same bank can be opened. ACTIVE command is used to open the rows and PRECHARGE (or the AUTO PRECHARGE hidden in the WRITE and READ commands, as used in this design) is used to close the rows. When issuing the commands for opening or closing the rows, both row address and bank address need to be provided.

For sequential access applications and those with page memory management, the proper address assignments and the use of the SDRAM pipeline feature deliver the highest performance SDRAM controller. However, this type of controller design is highly associated with the bus master cycle specification and will not fit the general applications. Therefore, this SDRAM controller design does not implement these custom features to achieve the highest performance through these techniques.

In this design, the ACTIVE command will be issued for each read or write access to open the row. After a tRCD delay is satisfied, READ or WRITE commands will be issued with a high sdr_A[10] to enable the AUTO REFRESH for closing the row after access. So, the clocks required for read/write cycle are fixed and the access can be random over the full address range. Read or write is determined by the sys_R_Wn status sampled at the rising edge of the clock before tRCD delay is satisfied. If a logic high is sampled, the state machine switches to c_READA. If a logic low is sampled, the state machine switches to c_WRITEA.For read cycles, the state machine switches from c_READA to c_cl for CAS latency, then switches to c_rdata for transferring data from SDRAM to bus master. The number of clocks the state machine stays in c_rdata state is determined by the burst length. After the data is transferred, it switches back to c_idle.For write cycles, the state machine switches from c_WRITEA to c_wdata for transferring data from bus master to SDRAM, then switches to c_tDAL. Similar to read, the number of clocks the state machine stays in c_wdata state is determined by the burst length. The time delay tDAL is the sum of WRITE recovery time tWR and the AUTO PRECHARGE timing delay tRP. After the clock rising edge of the last data in the burst sequence, no commands other than NOP can be issued to SDRAM before tDAL is satisfied.As mentioned in the INIT_FSM section above, the dash lines indicates possible state switching paths when tCK period is larger than timing delay spec.

e) Refresh cycle:

Similar to the other DRAMs, memory refresh is required. A SDRAM refresh request is generated by activating sdr_REF_REQ signal of the controller. The sdr_REF_ACK signal will acknowledge the recognition of sdr_REF_REQ and will be active throughout the whole refresh cycle. The sdr_REF_REQ signal must be maintained until the sdr_REF_ACK goes active in order to be recognized as a refresh cycle. Note that no system read/write access cycles are allowed when sdr_REF_ACK is active. All system interface cycles will be ignored during this period. The

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

sdr_REF_REQ signal assertion needs to be removed upon receipt of sdr_REF_ACK acknowledge, otherwise another refresh cycle will again be performed.

Upon receipt of sdr_REF_REQ assertion, the state machine CMD_FSM enters the c_AR state to issue an AUTO REFRESH command to the SDRAM. After tRFC time delay is satisfied, CMD_FSM returns to c_idle.

5. DATA PATH

The fig.11 shows the data path module with inputs and outputs as shown in the figure

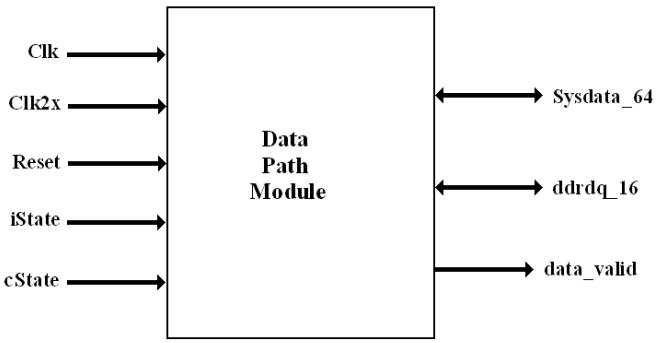


Figure .11.Data Path Module.

The data flow design between the SDRAM and the system interface. The module in this reference design interfaces between the SDRAM with 16-bit bidirectional data bus and the bus master with 64-bit bidirectional data bus. The user should be able to modify this module to customize to fit his/her system bus requirements. The data path module performs the data latching and dispatching of the data between the processor and DDR.

6. TIMING DIAGRAMS

The fig.12 and Fig.13 are the read cycle and write cycle timing diagrams of the reference design with the two CAS latency cycles and the burst length of four. In the example shown in the figures, the read cycle takes 10 clocks and the write cycle takes 9 clocks. The state variable c_State of CMD_FSM is also shown in these figures. Note that the ACTIVE, READ, WRITE commands are asserted one clock after the c_ACTIVE, c_READA, c_WRITEA states respectively. The values of the region filed with slashes in the system interface input signals of these figures are "don't care." For example, signal sys_R_Wn needs to be valid only at the clock before CMD_FSM switches to the c_READA or c_WRITEA states. Depending on the values of tRCD and tCK, this means the signal sys_R_Wn needs to be valid at state c_ACTIVE or the last clock of state c_tRCD.

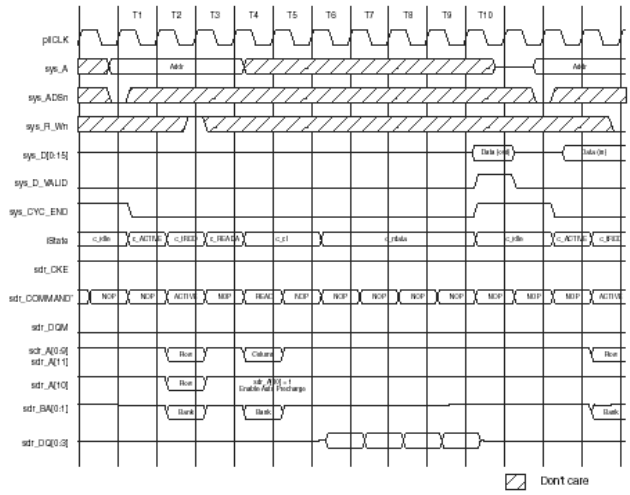


Figure .12 Read Cycle Timing Diagram

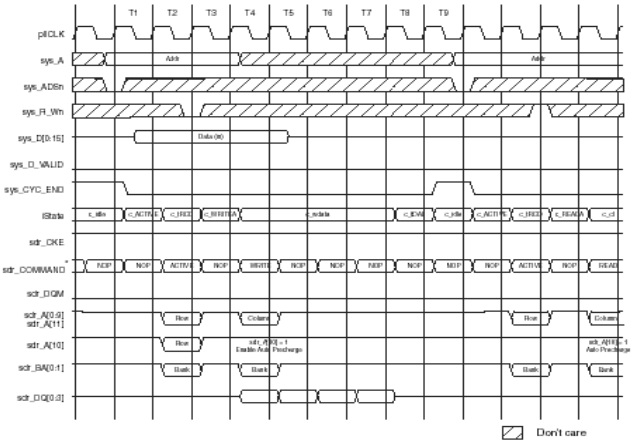


Figure.13 Write Cycle Timing Diagram.

The signal generation module generates the address and command signals required for DDR based on istate and cstate

6. IMPLEMENTATION AND RESULT

DUT mainly contains two parts, AHB slave and DDR SDRAM controller. The core is developed by using verilog [2]. In this work we have designed a High speed DDR SDRAM Controller with 64-bit data transfer which synchronizes the transfer of data between DDR RAM and External AHB compatible devices. This can be used in ARM based SOC design. The advantages of this controller compared to SDR SDRAM is that it synchronizes the data transfer, and the data transfer is twice as fast as previous, the production cost is also very low. This core is verified by using testbench and several testcases, which cover most of the functionality of the core. The simulations of specified functions were conducted by the Modelsim tool [3].The fig.14, Fig.15 and Fig 16 shows the waveforms initial, command and data path of the core. Fig.17 and 18 shows the RTL schematic of top module and controller respectively.

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

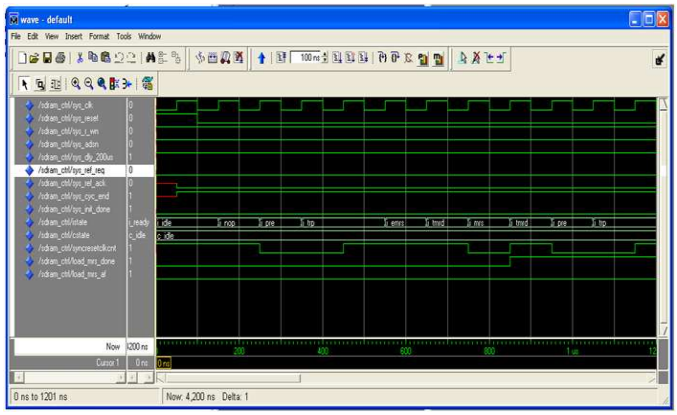


Figure.14 Waveforms of initial FSM

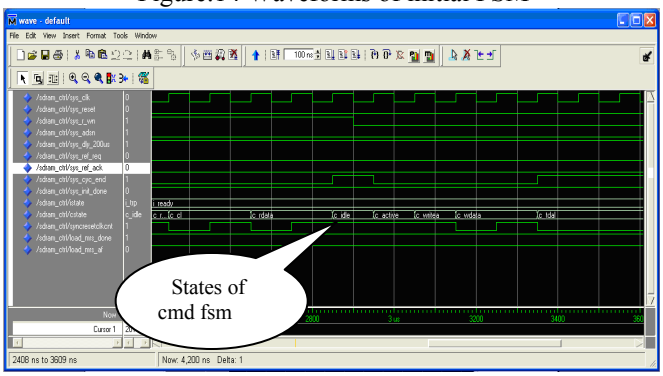


Figure.15 Waveforms of Command FSM

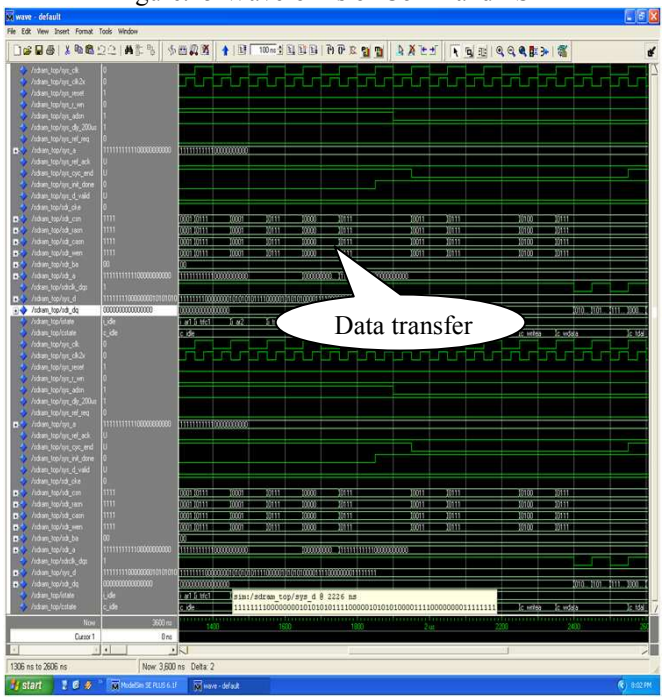


Figure .16 Waveforms Of Data path

For verification of core, two BFM's are used, AHB BFM at the input side and DDR memory model BFM at the output side. Core has been tested for main functionality of DDR SDRAM controller. AHB slave tested for both split and burst transfers. Our implementation of the design was analyzed by using ISE FPGA tool from Xilinx [1]. Designs are mapped on to Spartan FPGA. The whole system has been placed and routed in to the XC3S500e-5-PQ208 FPGA chip. The numbers of slices utilized by the design are 1340 out of 4656, 28% of the utilization.

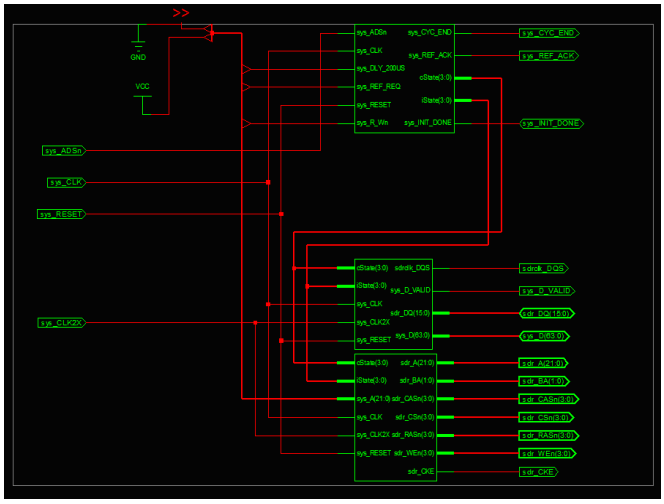


Figure .17.RTL Schematic of Top Module

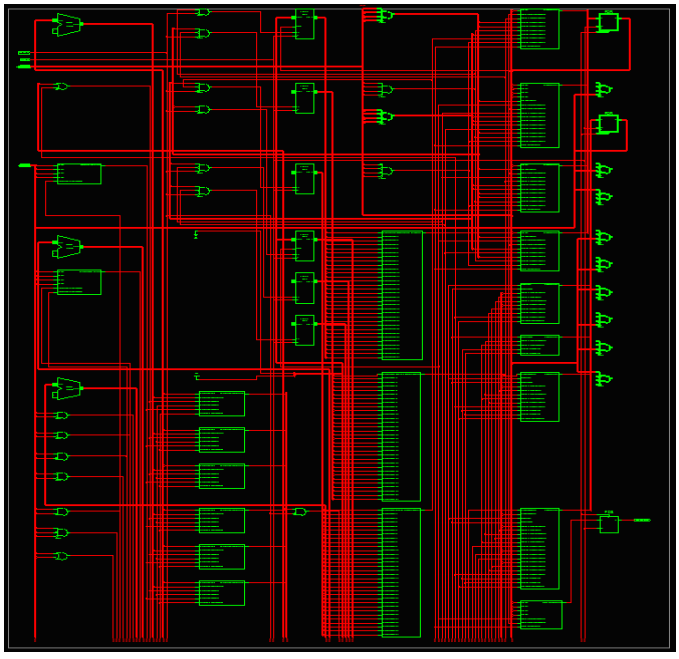


Figure .18.RTL Schematic of Controller

HDL Synthesis report of the core as shown below

Macro Statistics

# Adders/Subtractors	: 7
# 3-bit adder	: 2
#32-bit subtractor	: 6
# Counters	: 7
# 32-bit down counter	: 7

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

Registers : 18
1-bit register : 8
16-bit register : 2
#3-bit register : 1
32-bit register : 4
#4-bit register : 2
Latches : 2
64-bit latch : 1

Device utilization summary:

Selected Device : 3s400tq144-5

Number of Slices	:	482 out of 3584	14%
Number of Slice Flip Flops	:	468 out of 7168	6%
Number of 4 input LUTs	:	8944 out of 7168	12%
Number of IOs	:	165	
Number of bonded IOBs	:	165 out of 97	629% (*)
IOB Flip Flops	:	71	
Number of GCLKs	:	3 out of 8	37%

Timing Summary:

Speed Grade: -5

Minimum period: 9.802ns (Maximum Frequency: 102.021MHz)

Minimum input arrival time before clock: 5.954ns Maximum output required time after clock: 10.704ns

Maximum combinational path delay: 8.451ns

DDR2 SDRAM is the second generation of DDR SDRAM. DDR2 SDRAM improves on DDR SDRAM by using differential signaling and lower voltages to support significant performance advantages over DDR SDRAM. DDR SDRAM standards are still being developed and improved [20].

REFERENCES

[1]. http://www.xilinx.com
[2]. Samir Palnitkar, Pearson 2nd edition "Verilog HDL, A Guide to Digital Design and Synthesis.
[3]. http://www.micron.com
[4].Bandwidth, Area Efficient and Target Device Independent DDR SDRAM Controller by T. Mladenov, F. Mujahid, E. Jung, and D. Har , Proceedings Of World Academy Of Science, Engineering And Technology Volume 18 December 2006 Issn 1307-6884
[5].Application of DDR Controller for High-speed Data Acquisition Board by Zude Zhou, Songlin Cheng, and Quan Liu *School of Information Engineering, Wuhan University of Technology,* Proceedings of the First International Conference on Innovative Computing, Information and Control (ICICIC'06)IEEE
[6] An All-Digital Delay-Locked Loop for DDR SDRAM Controller Applications Ching-Che Chung, Pao-Lung Chen, and Chen-Yi Lee ,Dept. of Electronics Engineering / National Chiao Tung University OO ©2006 IEEEz[7] A Novel Design of DDR-based Data Acquisition Storage Module in a Digitizer by Jie Guo, Yibing Shi, Zhigang Wang School of Automation Engineering University of Electronic Science and Technology of China May 2, 2009 at 02:42 from IEEE Xplore.
[8] Double Data Rate (DDR) SDRAM Specification, JEDEC STANDARD, JESD79E, May 2005
[9] The Love/Hate Relationship with DDR SDRAM Controllers, Graham Allan, MOSAID
[10] 128Mb DDR SDRAM, Device Specification, Hynix, April 2006
[11] Altdq & Altdqs Megafunction, User Guide, Altera, March 2005
[12] PLLs in Stratix II & Stratix II GX Devices, April 2006
[13] How to Use DDR SDRAM, User's Manual, Document No. E0234E40, ALPIDA, September 2005
[14] Initialization Sequence for DDR SDRAM, Technical Note, TN-46-08, Micron.
[15] DDR SDRAM Controller, Reference Design RD1020, Lattice, Semiconductor Corporation, April 2004.
[16] DDR SDRAM Controller Using Virtex-4 FPGA Devices, Oliver Despaux, Application Note, March 27, 2006.
[19] ALTERA, *DDR SDRAM Controller White Paper, Ver1.1*, 2002, 8.
[20] Guo Li, Zhang Ying, Li Ning, and Guo Yang, "The Feature of DDR SDRAM and the Implementation of DDR SDRAM Controllers via VHDL", *The Journal of China Universities of Posts and Telecommunications*, 2002, vol.9, no. 1, pp. 61-65. Proceedings

AUTHORS PROFILE

[1] Dr. R. Shashikumar is presently working as a Professor in E & C dept, SJCIT, Chikballapur, Karnataka, India. He is having 10 years of teaching and 6 years of Industry experience. His areas of interest includes ASIC, FPGA, Network Security.

[2] Prof.C.N.Vijayakumar M.E, MISTE, MIE, MIETE is presently working as a HOD and Assistant Professor in the department of Telecommunication engg , SJCIT, Chikballapur, Karnataka, India. He is having 15 years of teaching experience. His areas of interest are Power Electronics, Low Power VLSI, ASIC and Control System.

[3]Mr. M.N.NagendraKumar M.E, MISTE is working as a Asst.Professor in the Dept of E & C, SJCIT, Chikballapur, Karnataka, India. He is having 13 years of teaching experience. His areas of interest are Power electronics, Control System and VLSI.

[4]Mr. C.S.Hemanthkumar M.Tech, MISTE is working as a Sr.Lecturer in the Dept of E&C, JVIT, Bangalore, India. His areas of interest are VLSI, Embedded, Signal Processing.

High Throughput of WiMAX MIMO-OFDM Including Adaptive Modulation and Coding

Hadj Zerrouki[1], Mohamed Feham[2]
Laboratoire de Systèmes de Technologies de l'Information
et de Communication (STIC)
University Abou Baker Belkaid, Tlemcen, Algeria.

***Abstract*— WiMAX technology is based on the IEEE 802.16 specification of which IEEE 802.16-2004 and 802.16e amendment are Physical (PHY) layer specifications. IEEE 802.16-2004 currently supports several multiple-antenna options including Space-Time Codes (STC), Multiple-Input Multiple-Output (MIMO) antenna systems and Adaptive Antenna Systems (AAS). The most recent WiMAX standard (802.16e) supports broadband applications to mobile terminals and laptops. Using Adaptive Modulation and Coding (AMC) we analyze the performance of OFDM physical layer in WiMAX based on the simulation results of Bit-Error-Rate (BER), and data throughput. The performance analysis of OFDM-PHY is done. In this paper, an extension to the basic SISO mode, a number of 2x2 MIMO extensions are analysed under different combinations of digital modulation (QPSK, 16-QAM and 64-QAM) and Convolutional Code (CC) with 1/2 , 2⁄3 and 3/4 rated codes. The intent of this paper is to provide an idea of the benefits of multiple antenna systems over single antenna systems in WiMAX type deployments.**

Keywords-WiMAX; MIMO; OFDM; AMC; Space Time Block Codes; Spatial Multiplexing

I. INTRODUCTION

The first WiMAX systems were based on the IEEE 802.16-2004 standard [1]. This targeted fixed broadband wireless applications via the installation of Customer Premises Equipment (CPE). In December, 2005 the IEEE completed the 802.16e-2005 [2] amendment, which added new features to support mobile applications. The resulting standard is commonly referred to as mobile WiMAX.

An ever crowded radio spectrum implies that future demands must be met using more data throughput wireless technologies. Since system bandwidth is limited and user demand continues to grow, spectral efficiency is vital. One way to improve link capacity, and potentially increase spectral efficiency, is the application of MIMO. It is well reported in the literature that MIMO physical (PHY) layer techniques have the potential to significantly increase bandwidth efficiency in a rich scattering environment [3]. Orthogonal Frequency Division Multiplexing (OFDM) is a well-established technique for achieving low-cost broadband wireless connectivity, and has been chosen as the air interface for a range of new standards, including IEEE802.16d/e. The ideas of MIMO and OFDM have been combined by a number of authors to form a new class of MIMO-OFDM system [4] [5]. This approach represents a promising candidate for WiMAX applications.

WiMAX standard supports a full-range of smart antenna techniques, including spatial transmit diversity and spatial multiplexing (SM). Spatial transmit diversity is achieved by applying Alamouti's Space-Time coding. SM can also be employed to increase the error-free peak throughput. Higher order modulation schemes with SM increase the link throughput, but require high SNR to achieve low Packet Error Rates (PER). Space-Time Block Coding (STBC) provides strong diversity gain, but cannot increase the link throughput without the use of Adaptive Modulation and Coding (AMC), and therefore AMC has become a standard approach in recently developed wireless standards, including WiMAX. The idea behind AMC is to dynamically adapt the modulation and coding scheme to the channel conditions so as to achieve the highest spectral efficiency at all times. Adaptive modulation changes the coding scheme and/or modulation method depending on channel-state information - choosing it in such a way that it squeezes the most out of what the channel can transmit.

This paper investigates the performance of the WiMAX standard when MIMO techniques are applied. Bit Error Rate (BER) and throughput results are presented for a MIMO with OFDM system that uses the coding and modulation schemes defined in the WiMAX standard IEEE 802.16-2004. Results are compared with basic SISO operation.

II. THE WIMAX PHY DESCRIPTION

The IEEE 802.16 standard was firstly designed to address communications with direct visibility in the frequency band from 10 to 66 GHz. Due to the fact that non-line-of-sight transmissions are difficult when communicating at high frequencies, the amendment 802.16a was specified for working in a lower frequency band, between 2 and 11 GHz. The IEEE 802.16d specification is a variation of the fixed standard (IEEE 802.16a) with the main advantage of optimizing the power consumption of the mobile devices. The last revision of this specification is better known as IEEE 802.16-2004 [1].

On the other hand, the IEEE 802.16e standard is an amendment to the 802.16-2004 base specification with the aim of targeting the mobile market by adding portability.

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

WiMAX standard-based products are designed to work not only with IEEE 802.16-2004 but also with the IEEE 802.16e specification. While the 802.16-2004 is primarily intended for stationary transmission, the 802.16e is oriented to both stationary and mobile deployments.

A. PHY Layer Overview

WiMAX is not truly new; rather, it is unique because it was designed from the ground up to deliver maximum throughput to maximum distance while offering 99.999 percent reliability. To achieve this, the designers (IEEE 802.16 Working Group D) relied on proven technologies for the PHY including orthogonal frequency division multiplexing (OFDM), time division duplex (TDD), frequency division duplex (FDD), Quadrature Phase Shift Keying (QPSK), and Quadrature Amplitude Modulation (QAM), to name only a few. WiMAX has a scalable physical-layer architecture that allows for the data rate to scale easily with available channel bandwidth. This scalability is supported in the OFDMA[1] mode, where the FFT (fast Fourier transform) size may be scaled based on the available channel bandwidth. For example, a WiMAX system may use 128, 512, or 1024 FFTs based on whether the channel bandwidth is 1.25MHz, 5MHz, or 10MHz respectively. This scaling may be done dynamically to support user roaming across different networks that may have different bandwidth allocations.

B. OFDM Parameters in WiMAX

As mentioned previously, the fixed and mobile versions of WiMAX have slightly different implementations of the OFDM physical layer. Fixed WiMAX, which is based on IEEE 802.16-2004, uses a 256 FFT-based OFDM physical layer. Mobile WiMAX, which is based on the IEEE 802.16e-2005[2] standard, uses a scalable OFDMA-based physical layer. In the case of mobile WiMAX, the FFT sizes can vary from 128 bits to 2048 bits.

Table I shows the OFDM-related parameters for both the OFDM-PHY and the OFDMA-PHY. The parameters are shown here for only a limited set of profiles that are likely to be deployed and do not constitute an exhaustive set of possible values.

C. WiMAX OFDMA-PHY

In Mobile WiMAX, the FFT size is scalable from 128 to 2048. Here, when the available bandwidth increases, the FFT size is also increased such that the subcarrier spacing is always 10.94 kHz. This keeps the OFDM symbol duration, which is the basic resource unit, fixed and therefore makes scaling have minimal impact on higher layers. A scalable design also keeps the costs low. The subcarrier spacing of 10.94 kHz was chosen as a good balance between satisfying the delay spread and Doppler spread requirements for operating in mixed fixed and mobile environments. This subcarrier spacing can support delay-spread values up to 20 µs and vehicular mobility up to 125 km/h when operating in 3.5GHz. A subcarrier spacing of 10.94 kHz implies that 128, 512, 1024, and 2048 FFT are used when the channel bandwidth is 1.25MHz, 5MHz, 10MHz, and 20MHz, respectively. It should, however, be noted that mobile WiMAX may also include additional bandwidth profiles.

TABLE I. OFDM PARAMETERS USED IN WIMAX

Parameter	Fixed WiMAX OFDM	Mobile WiMAX Scalable OFDMA[a]			
FFT size	256	128	**512**	1024	2048
Number of used data subcarriers	192	72	**360**	720	1440
Number of pilot subcarriers	8	12	**60**	120	240
Number of null/guardband subcarriers	56	44	**92**	184	360
Cyclic prefix or guard time	1/4, **1/8**, 1/16, 1/32				
Channel bandwidth (MHz)	3.5	1.25	**5**	10	20
Subcarrier frequency spacing (kHz)	15.625	**10.94**			
Useful symbol time (µs)	64	**91.4**			
Guard time assuming 12.5% (µs)	8	**11.4**			
OFDM symbol duration (µs)	72	**102.9**			
Number of OFDM symbols in 5 ms frame	69	**48.0**			

a. Boldfaced values correspond to the OFDMA parameters used in our evaluation of the **Fixed WiMAX** standard.

III. SYSTEM MODEL DESCRIPTION

Since increase the link throughput systems is the main goal of this work. Figure 1 depicts the block scheme of a typical structure of a system combining Space Time Coding (STC) with MIMO OFDM enabled WiMAX simulator used in this paper. The Block diagram represents the whole system model or the signal chain at base band. The block system is divided into 2 main sections namely the transmitter and the receiver.

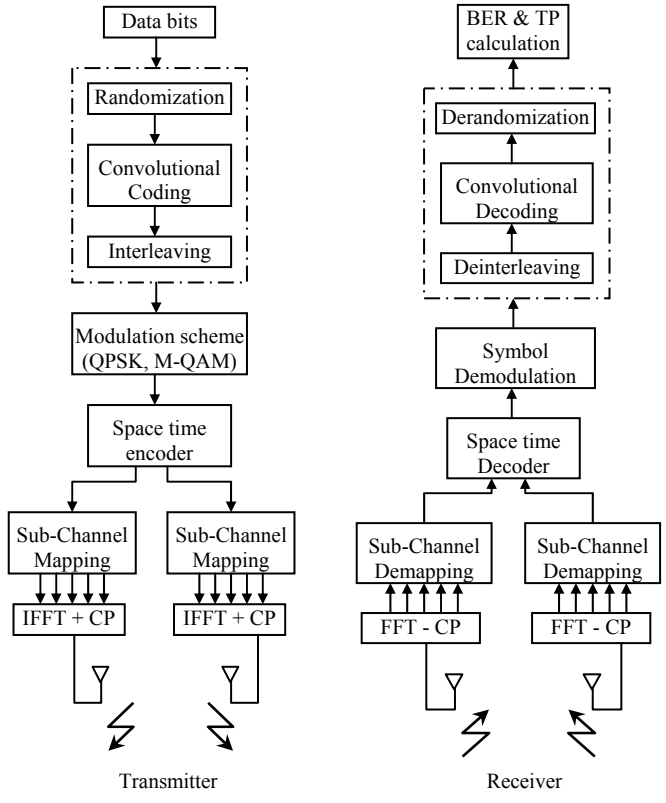


Figure 1. System block diagram for WiMAX MIMO OFDM simulator used in this paper.

[1] Orthogonal Frequency Division Multiplexing Access.
[2] Although the scalable OFDMA scheme is referred to as mobile WiMAX, it can be used in fixed, nomadic, and mobile applications.

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

A. Transmitter

The data is generated from a random source, consists of a series of ones and zeros. Since the transmission is done block wise, when forward error correction (FEC) is used, the size of the data generated depends on the block size used, modulation scheme used to map the bits to symbols (QPSK, *M*-QAM), and whether FEC is used or not [1]. The generated data is passed on to the next stage, either to the FEC block or directly to the symbol mapping if FEC is not used.

1) Channel coding: There are various combinations of modulations and code rates available in the OFDMA burst. Channel coding includes the randomization of data, forward error correction (FEC) encoding, interleaving, and modulation. In some cases, transmitted data may also be repeated on an adjacent subcarrier.

2) Randomization: Randomization of the data sequence is typically implemented to avoid the peak-to-average power ratio (PAPR) increasing beyond that of Gaussian noise, thus putting a boundary on the nonlinear distortion created in the transmitter's power amplifiers. It can also help minimize peaks in the spectral response.

3) Forward error correction (FEC): In our case, the error correcting codes are used, the data generated is randomized so as to avoid long run of zeros or ones, the result is ease in carrier recovery at the receiver. The randomized data is encoded using tail biting convolutional codes (CC) whose constraint length is 7 and the native code rate is ½ (puncturing of codes is provided in the standard to produce higher code rates). Finally interleaving is done by two stage permutation, first to avoid mapping of adjacent coded bits on adjacent subcarriers and the second permutation insures that adjacent coded bits are mapped alternately onto less or more significant bits of the constellation, thus avoiding long runs of lowly reliable bits.

4) Modulation: There are three modulation types available for modulating the data onto the subcarriers : QPSK, 16QAM, and 64QAM used with gray coding in the constellation map. In the Uplink, the transmit power is automatically adjusted when the modulation coding sequence (MCS) changes to maintain the required nominal carrier-to-noise ratio at the BS receiver. 64QAM is not mandatory for the Uplink.

5) Space Time Encoder (MIMO encoder): The Space Time Encoder stage converts one single input data stream into multiple output data streams. How the output streams are formatted depends on the type of MIMO method employed. Different symbols are simultaneously transmitted over these antennas to reduce noise interference. The receiver after receiving the signal retrieves the bits using Maximum Likelihood decoding algorithm and passes the data to the guard band removal block.

6) IFFT and cyclic prefixand: An'*N*' point inverse discrete fourier transform (IDFT) of '*X(k)*'is defined as:

$$x(n) = \frac{1}{N}\sum_{n=0}^{N-1} x(k) e^{j\frac{2\pi kn}{N}} \quad \text{for '}n\text{' =0,1,.... } N\text{-1.} \qquad (1)$$

From the equation we can infer that this is equivalent to generation of OFDM symbol. An efficient way of implementing IDFT is by inverse fast Fourier transform (IFFT). Hence IFFT is used in generation of OFDM symbol. The addition of cyclic prefix is done on the time domain symbol obtained after IFFT. The IFFT size ('*N*' value) is considered as 512 in simulations. This data is fed to the channel which represents 'Rayleigh fading channel model' and also implements multipath as shown in block diagram.

B. Receiver

The first thing done at receiver (in simulation) is removal of cyclic prefix, thus eliminating the inter symbol interference (ISI). Data is then passed through the serial to parallel converter of size 512 and then fed to the FFT for frequency domain transformation. The signal was distorted by the channel, to reconstruct the original signal we need information as to how the channel acted on the transmitted signal so that we can mitigate its effect. This is called equalization. In an OFDM system, this is done by channel estimation and interpolation, and reverse process (including deinterleaving and decoding) is executed to obtain the original data bits. As the deinterleaving process only changes the order of received data, the error probability is intact.

C. Rayleigh Fading Channel

Rayleigh Fading is one kind of statistical model which propagates the environment of radio signal. According to Rayleigh distribution magnitude of a signal which has passed though the communication channel and varies randomly. Rayleigh Fading works as a reasonable model when many objects in environment which scatter radio signal before arriving of receiver. When there is no propagation dominant during line of sight between transmitter and receiver on that time Rayleigh Fading is most applicable. On the other hand Rician Fading is more applicable than Rayleigh Fading when there is dominant line of sight. During our simulation we used Rayleigh Fading when we simulate the performance of BER and throughput Vs Signal to Noise Ratio.

IV. ENHANCEMENT WITH MIMO LINK ADAPTATION

It is well-understood that spectral efficiency is the key to good system design. Without loss of generality, the normalized effective system efficiency can be written as:

$$\xi_{sys} = \frac{system\ data\ throughput}{total\ radio\ resouces\ allocated} \qquad (2)$$

Two clear approaches emerge to improve the effective efficiency. Firstly the system data throughput can be improved by using methods such as high level AMC and MIMO. Secondly, improvements can be made to reduce the amount of radio resource required in the system. This section focuses on the system performance enhancements with MIMO Link Adaptation (LA) in combination with OFDMA. Using a

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

statistical approach, it is possible to demonstrate the potential benefits of a relay enhanced mobile WiMAX deployment.

A. Throughputs and coverages for WiMAX system

It is well-known that MIMO promises transmission efficiency enhancement which achieves one aspect of efficiency enhancement. IEEE 802.16e standard supports a full range of smart antenna technologies. Together with modulation and coding, the link throughput for each user can be calculated from Packet Error Rate (*PER*) by:

$$C_{link} = \frac{N_D N_b R_{FEC} R_{STC}}{T_s} \times (1 - PER) \quad (3)$$

Where, T_s, N_D, N_b, R_{FEC} and R_{STC} denote the OFDMA symbol duration, the number of assigned data subcarriers, the number of bits per subcarrier, FEC coding rate, and space-time coding rate for the user. Equation (3) implies that, a combination of MIMO, AMC and flexible sub-channelization is required to maximize the link performance.

B. MIMO scenarios description

1) Space-Time Block Coding (STBC): Our Fixed WIMAX simulator implements the Alamouti scheme [7] on the Downlink to provide transmit and receive diversity. This scheme uses a transmission matrix [s_1, $-s_2^*$; s_2, s_1^*], where s_1 and s_2 represents two consecutive OFDMA symbols.

2) Spatial Multiplexing (SM): WiMAX system supports SM to increase the peak error free data rate [8]. The idea behind spatial multiplexing is that multiple independent streams can be transmitted in parallel over multiple antennas and can be separated at the receiver using multiple receive chains through appropriate signal processing. This can be done as long as the multipath channels as seen by the various antennas are sufficiently decorrelated, as would be the case in a scattering-rich environment. Spatial multiplexing provides data rate and capacity gains proportional to the number of antennas used, a 2x2 SM system can double the peak data rate. This comes at the expense of sacrificing diversity gain, and hence a much higher SNR is required.

3) Adaptive Modulation and Coding: WiMAX systems use adaptive modulation and coding in order to take advantage of fluctuations in the channel. The basic idea is quite simple: Transmit as high a data rate as possible when the channel is good, and transmit at a lower rate when the channel is poor, in order to avoid excessive dropped packets. Lower data rates are achieved by using a small constellation, such as QPSK, and low-rate error-correcting codes, such as rate convolutional or turbo codes. The higher data rates are achieved with large constellations, such as 64 QAM, and less robust error correcting codes; for example, rate convolutional, turbo, or LDPC codes. In all, 52 configurations of modulation order and coding types and rates are possible, although most implementations of WiMAX offer only a fraction of these[9].

In our case, by using six (6) of the common WiMAX burst profiles; it is possible to achieve a large range of spectral efficiencies. This allows the throughput to increase as the signal-to-interference-plus-noise ratio (SINR) increases following the trend promised by Shannon's formula. In this case, the lowest offered data rate is QPSK and rate 1/2 Convolutional codes; the highest data-rate burst profile is with 64 QAM and rate 3/4 Convolutional codes. The achieved throughput normalized by the bandwidth is defined as:

$$T = (1 - BLER)_r \log_2(M) \quad bps / Hz \quad (4)$$

where *BLER* is the block error rate, $r \leq 1$ is the coding rate, and *M* is the number of points in the constellation. For example, 64 QAM with rate 3/4 codes achieves a maximum throughput, when BLER→0; QPSK with rate 1/2 codes achieves a best-case throughput.

V. Simulation Results

In this section SISO and MIMO BER and Throughput results are presented using the Fixed WiMAX simulator. The Simulation model was implemented in Matlab® 7. The PHY parameters used in simulation are given in Table I, with boldfaced values correspond to the OFDMA-PHY parameters. A carrier frequency of 2GHz is considered. For Spatial Multiplexing, an MMSE receiver is used to remove the inter-stream interference on a per sub-carrier basis. The link throughput is calculated from the PER as given by equation (3) in section IV. On the downlink, we consider that no sharing of OFDMA symbol. In this case of single-user MIMO, the multiple streams are intended for the same receiver; we have considered transmission formats with only a single stream for a single user as a basic OFDM in fixed WiMAX.

Figure 2 and 3 displays the performance of SISO and 2x2 MIMO STBC systems. For comparison purposes, the SNR versus BER graph is plotted for different modulation schemas and Convolutional Coding (CC) rates are showed.

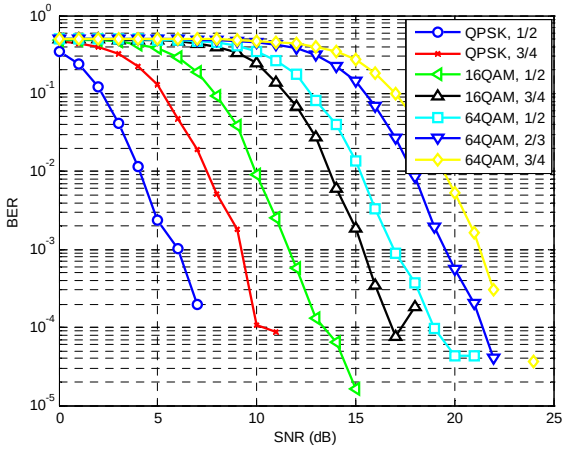


Figure 2. BER performance of SISO WiMAX system.

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

It is evident from the graph that the QPSK modulation scheme with 1/2 CC rate is better suited for OFDM transmission in terms of BER performance, and it can be see that the BER performance MIMO system with 2x2 STBC yields a gain of about 3 dB over the corresponding SISO system at a BER of 10^{-3}.

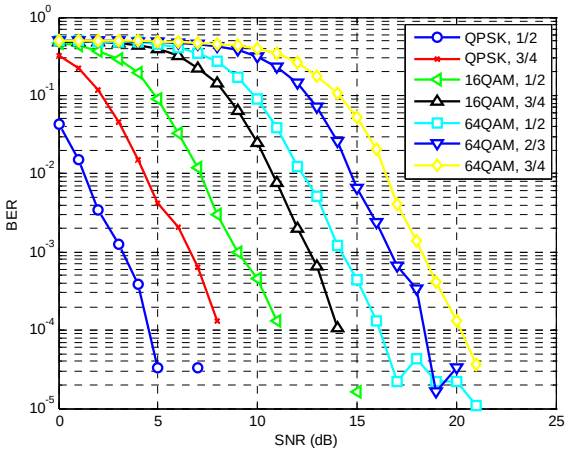


Figure 3. BER performance of 2x2 STBC WiMAX system.

Figure 4 and 5 present the throughput versus SNR graphs for the SISO and MIMO 2x2 STBC scenarios respectively. We observe that STBC offers a significant performance gain of 3dB to 4dB, the exact value depend on the selected link-speed. As we see here, STBC does not improve the data throughput, however at a given SNR STBC can provide a significant increase in throughput when combined with suitable link adaptation, since higher throughput modes can be used at much lower values of SNR. STBC increases the robustness of the WiMAX system by coding over the different transmitter branches and over space and temporal dimension. In this way, the spectral efficiency of MIMO is exploited by adding extra redundancy to improve the performance.

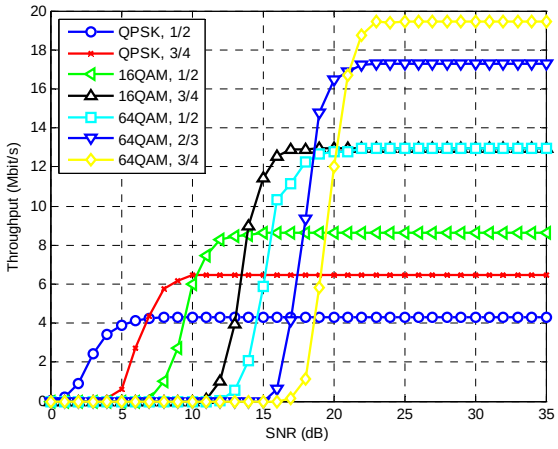


Figure 4. Throughput of SISO WiMAX system.

Figure 6 illustrates the simulated MIMO 2x2 SM system throughputs versus SNR. The main advantage of SM is that it directly exploits the MIMO channel capacity to improve the data throughput by simultaneously transmit different signals on different transmit antennas, at the same carrier frequency. The main disadvantage is that no redundancy is added and, thus, it might suffer from poor link reliability. To overcome this problem additional channel coding can introduced. This, however, reduces its data rate advantage. As expected, the SM 2x2 modes doubles the peak error-free throughput of every link-speed. However, at low SNR values the throughput of SM is less than STBC.

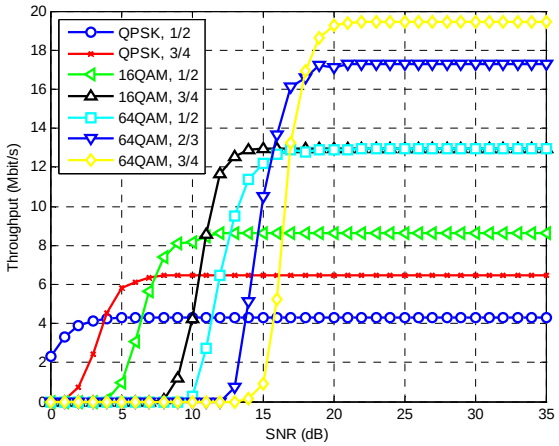


Figure 5. Throughput of 2x2 STBC WiMAX system.

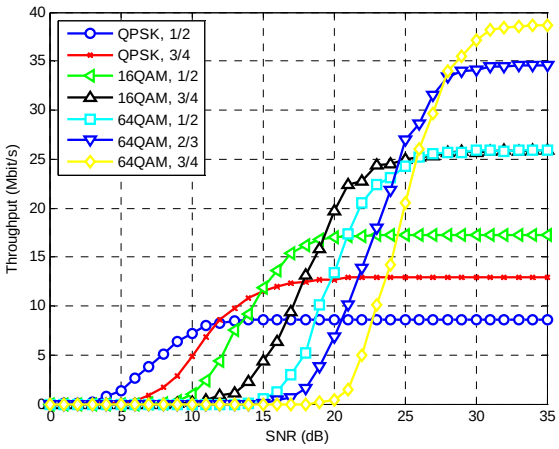


Figure 6. Throughput of 2x2 SM WiMAX system.

Figure 7 shows the throughput envelope versus SNR for all the investigated fixed WiMAX scenarios: SISO, 2x2 STBC, and 2x2 SM. This envelope assumes the use of adaptive modulation and coding (AMC) to maximize the expected throughput. Obviously, both MIMO schemes outperform the SISO scenario. However, for a very spatially correlated channel, the SM method can be worse than SISO. In this case STBC performance would tend to that of SISO. The STBC produces the best performance at low to medium values of SNR, due to its robustness in poor channel conditions. On the

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

other hand, at high SNR the increased error-free data rate makes SM the best choice.

WiMAX system supports Adaptive MIMO Switching (AMS) to select the best MIMO scheme. Figure 7 clearly shows that for the channel conditions analysed here, the switching point between STBC and SM is 20dB. This value will increase with increasing spatial correlation.

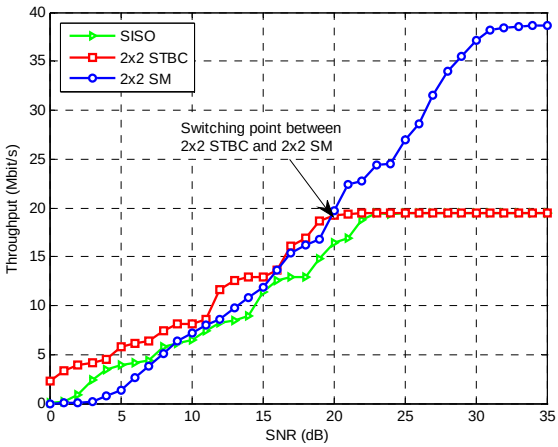


Figure 7. Adaptive Modulation and Coding (AMC) and switching point between STBC 2x2 and SM 2x2.

VI. CONCLUSION

In this paper, we considered the calculation of data throughput of the downlink of OFDMA-based IEEE802.16 WiMAX systems; this paper has presented a detailed study of the benefit of MIMO when applied to WiMAX. Throughput results were presented for diverse scenarios. At lower values of SNR STBC is preferred. However, at high SNR Adaptive MIMO Switching should be used to switch to SM. Give that SM 2x2 doubles the error-free throughput, at high SNR this scheme leads to the highest throughput. In practice, the viability of SM (and the value of the SNR switching threshold) depends on the level of spatial correlation.

REFERENCES

[1] IEEE Std 802.16TM-2004, "Part 16: Air interface for fixed broadband wireless access systems," Oct 2004.

[2] IEEE Std 802.16Etm-2005, "Part 16: Air interface for fixed and mobile broadband wireless access systems," Feb. 2006.

[3] A. Paulraj R. Nabar and D. Gore, Introduction to space-time wireless communications. Cambridge university press, 2003.

[4] H. Sampath, S. TAlwar, J. Tellado, V. Erceg and A. Paulraj, "A forth-generation MIMO-OFDM broadband wireless system: design performance, and field trial results," IEEE Communications Magazine, Vol. 40, No. 9, pp 143-149, Sep. 2002.

[5] A. Zelst and T. Schenk, "Implementation of a MIMO OFDM-based wireless LAN system," IEEE Trans. Axoustics, Speech and Signal Processing, Vol. 52, No. 2, pp 483-494, Feb. 2004.

[6] C. Tarhini, T. Chahed, On capacity of OFDMA-based IEEE802.16 WiMAX including Adaptive Modulation and Coding (AMC) and intercell interference, LANMAN'2007, Princeton NJ, June 2007.

[7] M. Alamouti, "A simple transmit diversity technique for wireless communications," IEEE JSAC, Vol. 16, No. 8, Oct. 1998.

[8] G. J. Foschini, "Layered Space-Time Architecture for Wireless Communication in a Fading Environment when Using Multi-element Antennas," Bell Labs Tech. J. pp. 41-59, Autumn 1996.

[9] J. G. Andrews, A. Ghosh, R. Muhamed, "Fundamentals of WiMAX: Understanding Broadband Wireless Networking, " Prentice Hall PTR, Feb, 2007.

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

Performance Modeling and evaluation of Traffic management for Mobile Networks by SINR Prediction

K.K.Guatam
Department of Computer Science & Engineering
Roorkee Engineering & Management Technology Institute
Shamli(247774) India

Anurag Rai
Department of Information Technology
College of Engineering Roorkee
Roorkee (247667) India

***Abstract*— Over the recent years a considerable amount of effort has been devoted towards the performance evaluation and prediction of Mobile Networks. Performance modeling and evaluation of mobile networks are very important in view of their ever expending usage and the multiplicity of their component parts together with the complexity of their functioning. The present paper addresses current issues in traffic management and congestion control by (singal-to-interference-plus-noise ratio) SINR prediction congestion control, routing and optimization of cellular mobile networks.**

Keywords- Mobile Networks, Modeling, call admission control, QoS (Quality of Service) SINR.

I. INTRODUCTION

Over the recent years a considerable amount of effort has been devoted towards the performance, evaluation of wireless mobile networks (WMN). A considerable amount of research has been used to characterize user and calling behavior and their performance impact on wireless mobile networks. At present the mobility in most mobile networks is confined to the end users only.

With the development of mobile compression, the CAC schemes are generally adopted by setting thresholds for hand off calls and new call differently given the traffic condition and it is the maximum number of users that can be supported.[1,2] In realistic systems, information about the traffic management quality cannot; be instantaneous but is outdated to some degree. Firstly, the SINR estimation in the receiver takes some time and secondly, the user has to wait for the next channel allocation to report his SINR to the base station. The system may need to block incoming users if all of the entire band width has been used up to provide the highest QoS to existing users. However, if these existing users can be degraded to a lower but acceptable QoS level, it is possible to reduce the blocking probability without degrading the QoS of existing users. A graceful degradation mechanism is proposed in (1). Thus a system could free some bandwidth allocation for new users. In this paper, we address current issues in traffic management of cellular mobile networks by the use of SINR prediction, the SINR is calculated for mobile user equipments in every transmission time interval for the traffic management of mobile networks. In traffic management and congestion control, courcoubetis and series device new procedures and tools for the analysis of network traffic measurement.

II. MODEL DESCRIPTION

We consider uplink communication in a wireless mobile network. As an accepted call does not always send data frames. then for best traffic, we consider the activity factor ℓ as the probability that a call is active. We represent QoS requirement of traffic by required transmission rate. The required transmission rate can be obtained by setting the target level.

Often these intra–and inter–traffic interferences of calls can be large so that the target bit error rate of traffic interferences(BERIT) cannot be achieved temporarily, which is called outage.[3] The outage probability needs to approach zero (as close as possible) and can be different for each class. Here we assume for traffic management the allowed outage probability is the same.

III. OUTAGE PROBABILITY FOR TRAFFIC

In a mobile network, a traffic management that supports a single class of calls, the outage probability is given by [2,4].

$$Pout = Pr\ \{N^a + M^a > (3/2)\ G(x^{-1} - (Yb/N0)^{1})^{+1}\} \quad \ldots\ldots(1)$$

When N^a, M^a, G, X, Y^b, and N0 represent the number of active calls in the current call. Similarly, in a network that support L-Class of calls, we obtain

$$P^j out = Pr\ \{ \sum_{i=1}^{L} (Ybi / Ybj)\ Ci\ (Ni^a + Mi^a) \geq Aj\} = Pr$$

$$\{ \sum_{i=1}^{L} \theta I\ (Ni^a + Mi^a) \geq \eta j\} \quad \ldots\ldots\ldots\ldots (2)$$

When i, j represent traffic call classes (TCC), Ci is the number of orthogonal codes needs for a TCC, 'i'. By the

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

Gaussian random variable from the limit theorem and we can write control the outage probability of a TCC' 'j', as

$P^{j}out = Q(\eta j - \lambda / \partial\lambda)$……..(3)

Where $Q(\xi) = 1/\sqrt{2\pi} \int_{\xi}^{\infty} e^{-x2/2}\, dx$

And represent the total traffic receive single power (TRSP) i.e.

$\sum_{i=1}^{L} \theta I (Ni^{a}+Mi^{a})$

Therefore $\overline{\lambda} = (1+f1) \sum_{i=1}^{L} \theta I \overline{N}^{a}i,$…… (4)

And $\partial^{2}\lambda = \sum_{i=1}^{L} \theta^{2} I (\partial i^{2}+f2 \overline{N}^{a}i)$

Where $\overline{N}^{a}i$ and $\partial^{2}I$ indicate the mean and variance of $N^{a}I$.

According to the assumption of TCC equal outage probability for each class, $\eta I = \eta j$ for all I and j, therefore TCC received single power meets the following relation.

$\theta I / \theta j = CiXi (3G - 2CjXj) \diagup CjXj(3G+ 2CiXj)$…....(6)

This indicates that the power allocation refers to the target of TCC outage probability Call Admission Control (CAC):

IV. SYSTEM MODEL

The Communication system under consideration can be defined as $r[k] = \sum_{i=0}^{L} h[l]\& [k-l] + z[k]$ …….(7)

Where r [k] received call sequence h [l] unknown channel for traffic with memory L', z[k] is an independent and identically distributed Gaussian notice sequence. [5,6]

Then traffic management symbol sequence s [k] is drawn from M-ary alphabet, A with equal probability, the vector version of (1) can be written as

$$\begin{bmatrix} r[k] \\ : \\ : \\ r[1] \\ r[0] \end{bmatrix} = \begin{bmatrix} S[K-L] & \ldots & S[K] \\ : & \ldots & : \\ : & \ldots & : \\ S[1-L] & \ldots & S[1] \\ S[-L] & \ldots & S[0] \end{bmatrix} .. \begin{bmatrix} h[l] \\ : \\ : \\ h[1] \\ h[0] \end{bmatrix} + \begin{bmatrix} z[k] \\ : \\ : \\ z[1] \\ z[0] \end{bmatrix}$$

Where Sk is toeplitz data matrix.

V. CALL ADMISSION CONTROL FOR TRAFFIC MANAGEMENT

I call Admission Control for traffic Management, (CACTM) the outage Probability is very small, defined as $\frac{\partial Q(\eta)}{\partial \eta} < 0$ we can show that $\frac{\partial P_{out}}{\partial N_i} = \alpha_i \frac{\partial P_{out}}{\partial \overline{N_i}} > 0$

where α_i is the active factor for (CACTM) a class I call. It is clear that the average rate for mobile network (ARRMN) and outage probability increase with the number of users.

VI. NUMERICAL RESULT

We now compare the performance of the two CACs through numerical analysis. [7] The system bandwidth is 2.50MHz and each code can carry information bits at the rate of 19.2(kbps) so that the processing gain is 256. Two types of calls are considered to manifest the effect of traffic parameters on performance. Class 1 and 2 calls are voice traffic and we set their transmission rates after channel coding at 19.29(kbps). They have different Mobile Network Average Revenue Rate (MNARR) for the traffic management requirement of less than 10^{-4} and 10^{-6}, respectively, and their activity factors are set at 1.0. The coefficient for intercall interference modeling are chosen as f1 = 0.114 and f2 = 0.44(12).

VII. FRAME WORK

Angle – SINR Table:-

In order to make the directional routing effective for call admission control system, a node should know how to set its transmission direction effectively to transmit a packet to its neighbors. So each node periodically collects its neighborhood information and forms an Angle- SINR Table (AST). $Sinu^{s}m(t)$ (Signal – to – Interference and Noise Ratio) is a number associated with each link 1^{u} n, m, and is a measurable indicator of the strength of radio connection from node n to node m at an angle u with respect to n and as perceived by m at any point of time t for call admission control. AST of node n specifies the strength of radio connection of its neighbors with respect to n at a particular direction for call admission control. Angle - SINR Table for node n time t is shown below (Table I) where. we assume that nodes I, j and k are the neighbors of n. [9,10,11,12]

TABLE I. ANGLE – SINR TABLE (AST) FOR NODE n

Azimuth Angle (degree)	***SINR value as perceived by neighbors of rooters n at different angle w.r.t rooters n***		
	i	**j**	**K**
0	$SINR^{0}n,i^{(t)}$	$SINR^{0}n,j^{(t)}$	$SINR^{0}n,j^{(t)}$
30	$SINR^{30}n,i^{(t)}$	$SINR^{30}n,j^{(t)}$	$SINR^{30}n,j^{(t)}$
60	$SINR^{60}n,i^{(t)}$	$SINR^{60}n,j^{(t)}$	$SINR^{60}n,j^{(t)}$
…	…..	…..	….
330	$SINR^{330}n,i^{(t)}$	$SINR^{330}n,j^{(t)}$	$SINR^{330}n,j^{(t)}$
360	$SINR^{360}n,i^{(t)}$	$SINR^{360}n,j^{(t)}$	$SINR^{360}n,j^{(t)}$

In order to form ANGLE – SINR TABLE (AST), each node periodically sends a directional request in the form of a directional broadcast for the call admission control, sequentially in all direction. In this work, it has been done at 30 degree interval, covering the entire 360 degree space

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

sequentially. A node is i in the neighborhood of n will wait until it receives all the request packets generated by n in all direction in that occasion. In other words, node I accumulates the entire column of the AST of n for node I, I accumulates the entire column of the AST of n for rooters i. Here, rooters i, after receiving the first request from n, has to wait a pre-specified amount of time to make sure that the directional broadcasts by n in all direction are over. Rooters I sends this information from all the neighbors of n, the Angle-SINR Table of n would be complete.[13]

CONCLUSION

In this paper, we consider Call Admission Control for Traffic Management (CACTM) in Mobile Networks. Through the mathematical analysis and also present outage probability and a system model's for CAC. We also present an example for Call Admission Control for Traffic Management (CACTM) and present a frame work for the set up-the call admission control.

ACKNOWLEDGMENT

The author would like to thank Dr. H.N. Dutta, Director, REMTech for his moral support in carrying out the work.

REFERENCES

[1] J. Zhang , J. W. Mark, and S.Xuemin, "An adaptive handoff priority scheme for wireless MC-CDMA cellular networks supporting multimedia application ," in Proc. IEEE Globecom, Nov/Dec. 2004, pp. 3088-3092.

[2] Z. Liu and M. E. Zarki, "SIR – based call admission control for DSCDMA cellular system," IEEE J. Select. Areas Commun, vol . 12, May 1994, pp.638-644,

[3] R. j. Boucherie and Nico M. Van Dijk, "On a queueing network model cellular mobile telecommunications networks, "Operations Research, vol 48, no, 2000, pp . 38—49,

[4] X.Chao and W. Li, "Performance analysis of a cellular network with multiple classes of calls," IEEE Trans. Coomun, vol. 53, no. 9, pp. 1542-1550,2005.

[5] C. T. Chou and K.G. Shin, " Analysis of adaptive bandwidth allocation in wireless networks with multilevel degradable quality of service," IEEE Trans. Mobile Compute, vol. 3 no. 1, 2004, pp. 5-17,

[6] Maruf Mohammad , William Tranter " Blind Acquisition of short Burst with Per – Survivor Processing" IEEE TRANSCATION ON WIRELESS COMMUNICATION, vol. 6. No. 2. February 2007.

[7] Wei Li , and Xiulichao " Call admission Control for an adaptive Heterogeneous Multimedia Mobile Network" IEEE TRANSCATION ON WIRELESS COMMUNICATION, vol. 6. no. 2. February 2007. page no. 515-525.

[8] Jin-Cho Choi, Yound-June Choi, and Saewoong Bank "power – Based Admission Control for Multi Class Calls in QoS –Sensitive CDMA NETWORKS" IEEE TRANSCATIONS ON Wireless communication, vol, 6 no. 2 February 2007 page no. 469-472.

[9] Markus Jordan Gerd Ascheid and Heinrich Meyr , "Performance Evaluation of Opportunistic Beamforming with SINR Prediction for HSDPA", NEW Jersy 1993.

[10] M. Kawai, M.Nozaki and K.Gyoda, "A wireless Ad Hoc Community Network with Recongfigurable Topology Architecture", Proc.of the GLOBECOMM'98.1998.

[11] T. Ohira and K.Gyoda , "Electronically Steerable Passive Array Radiator (ESPAR),"Antennas for Law –cost Adaptive Beam Forming", IEEE International Conference on Phased Array System , dana Point , CA May 2000.

[12] S. Mandyopadhyay, K. Gyoda K. Hasuike,S.Horisawa,Y.Kado And s.Tawara, "An Adaptive MAC Protocol for wireless AD Hoc Community Network (WACNet) Using Electronically Steerable Passive Array Radiator Anteena ", submitted to GLIBECOM 2001.

[13] Fan Bai, N Sadagopan and A Helmy , " IMPROTANT : a Framework to systematically analyze the impact of mobility on performance of Routing Protocol for AdHoc Neworks", IEEE 2003

AUTHORS PROFILE

Authors Profile .. K K Gautam is the Dean in the Roorkee Engineering & Management Technology Institute, Shamli-247 774, India.

Anurag Rai is the Head of the Information Technology Department in College of Engineering Roorkee,Roorkee-247 667, India.

THAI RHETORICAL STRUCTURE ANALYSIS

Somnuk Sinthupoun
Department of Computer Science
Maejo University
Chiangmai, Thailand 50290

Ohm Sornil
Department of Computer Science
National Institute of Development Administration
Bangkok, Thailand 10240y

Abstract— **A rhetorical structure tree (RS tree) is a representation of discourse relations among elementary discourse units (EDUs). A RS tree is very useful to many text processing tasks employing relationships among EDUs such as text understanding, summarization, and question-answering. Thai language with its unique linguistic characteristics requires a unique RS tree construction technique. This paper proposes an approach for Thai RS tree construction which consists of three major steps: EDU segmentation, Thai RS tree construction, and discourse relation (DR) identification. Two hidden markov models derived from grammatical rules are used to segment EDUs, a clustering technique with its similarity measure derived from Thai semantic rules is used to construct a Thai RS tree, and a decision tree whose features extracted from the rules is used to determine the DR between EDUs. The proposed technique is evaluated using three Thai corpora. The results show the Thai RS tree construction and the DR identification effectiveness of 94.90% and 82.81%, respectively.**

Keywords- **Thai Language, Element Discourse Unit, Rhetorical Structure Tree, Discourse Relation.**

I. INTRODUCTION

A RS tree is a tree-like representation of discourse relations among elementary discourse units (EDUs) which can be defined as follows: RS tree = (Status, DR, Promotion, Left, Right) where Status is either nucleus or satellite EDU (nucleus expresses what is more essential to writer's purpose than satellite); DR is a Discourse Relation (DR); Promotion is a subset of EDUs; and Left, Right can be either NULL or recursively defined objects of type RS tree [4, 6].

Some researchers consider an EDU to be a clause or clause-like [6] excerpt, while others consider them to be a sentence [10] in discourse parsing. A number of techniques are proposed to determine EDU boundaries for English such as using discourse cues [5, 8, 9], punctuation marks [6, 9], and syntactic information [6, 10, 12].

There are many DRs. Some have a single nucleus such as elaboration and condition, while others have multiple nucleuses such as contrast [20]. A number of techniques for determining the DR between EDUs are proposed, such as using verb semantics [13] to build verb based events which represent EDUs, using cue phrase/discourse marker (e.g., "because", "however") [5], and using machine learning techniques [6].

For Thai RS tree construction, Sukvaree, et.al. [16] purposes a technique to construct a tree by using a spanning tree which make decision by discourse marker and focus of EDU, into two phases which consist of local and global EDU spanning tree. Using spanning tree, RS tree construction used the right adjacent rule to add right hand side to attach left hand side EDU.

For Thai DR recognition, Sukvaree, et.al. [16] purposes a technique to recognize a DR by using DR marker tag to recognize DR. If DR marker tag has many semantic DRs, the maximum probability value in chain of DR marker is used to identify the DR. There are four relations: cause-result, constion, contrast and elaboration. Wattanamethanont, et.al. [11] purposes a technique to recognize a DR by using Naïve bayes classifier. The feature of the machine learning are DR marker, key phrase and word co-occurrence. There are three relation: elaboration, logical and sequence.

This article proposes a new approach for Thai RS Tree construction which consists of two major steps: EDU segmentation and Thai RS tree construction. Two hidden markov models taking into account syntactic properties of Thai language are employed to segment EDUs, and a clustering technique with its similarity measure derived from linguistic properties of Thai language is employed to construct a Thai RS tree. Once the tree is created, DR between EDUs is then determined by a decision tree.

II. ISSUES IN THAI RS TREE

Thai language possesses unique characteristics syntactically and semantically. This makes techniques proposed for other languages not directly applicable to Thai language. A number of important issues are discussed in this section.

A. *Issues in Thai EDU Extraction*

No Explicit EDU boundary Thai language has no punctuation masks (comma, semi-colon and blank) and no special symbols (full stop) to identify the start and the end of EDUs. Unlike English contains specific symbols (e.g. '.', comma, semi-colon and blank) to specify the start and the end of EDUs, they can be used to separate a text into EDUs.

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

Therefore, EDU identification becomes a problem for Thai RS tree construction.

EDU1 EDU2 EDU3

Thai : $[w_1w_2\ldots w_mw_{m+1}w_{m+2}\ldots w_nw_{n+1}w_{n+2}\ldots w_o]$
English : $[w_1\ w_2\ \ldots\ w_m],[w_{m+1}\ w_{m+2}\ \ldots\ w_n];[w_{n+1}\ w_{n+2}\ \ldots\ w_o]$.
where w_i is a word in text.

Omission Problem Given two EDUs, an absence of subject, object or conjunction in anaphoric EDU may happen. Such as anaphoric EDU omits the subject that refers back to the object of antecedent EDU. Accordingly, EDU segmentation is ambiguous.

Thai : "เพื่อนจะขอยืมหนังสือ เพราะหาซื้อไม่ได้" (A friend's going to borrow this book. Because she hasn't been able to buy it.)

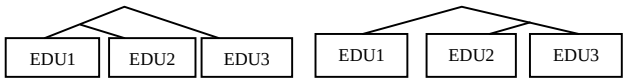
There are three Possible :
1) $[S(เพื่อน)V(จะขอยืม)O(หนังสือ)]_{EDU1}$ $[Because\ S(\Phi)\ V(หาซื้อไม่ได้)]_{EDU2}$
2) $[S(เพื่อน)V(จะขอยืม)O(หนังสือ)]_{EDU1}$ $[Because(\Phi)S(\Phi)V(หาซื้อไม่ได้)]_{EDU2}$
3) $[S(เพื่อน)V(จะขอยืม)O(\Phi)]_{EDU1}$ $[Because(\Phi)S(หนังสือ)V(หาซื้อไม่ได้)]_{EDU2}$

B. *Issues in Rhetorical Structure Tree Construction*

After EDUs are extracted correctly, relationships among EDUs need to be determined, and a number of issues need to be considered.

Adjacent Marker Problem Given three EDUs and two markers indicating different relationship, as shown in Ex. 1, there are two possible for the RS Tree. First, relationship between EDU1 and EDU2 is determined by using discourse marker "แต่" (but), next that between (EDU1, EDU2) and EDU3 is determined. On the other hand, the relationship between EDU2 and EDU3 is determined first by using discourse marker 'ถ้า' (if), next that between (EDU2, EDU3) and EDU1 is determined.

Ex. 1. EDU1: ศาลได้มีคำสั่งให้แยกสินสมรส (A court has order to separate marriage property.)
EDU2: แต่ถ้าภริยาหรือสามีคัดค้าน (But if a wife or a husband disagree,)
EDU3: ศาลจะสั่งยกเลิกการแยกได้ (a court will have order to cancel separation.)



a) A RS Tree (BUT) b) A RS Tree (IF)

Fig. 1. Adjacent marker problem

Omission Problem The absence of Subject, Object or Preposition which is modifier nucleus of VP especially in anaphoric EDU of Thai language are often occur.

Ex. 2. EDU1: ศาลได้มีคำสั่งให้แยกสินสมรส (A court has order to separate marriage property.)
EDU2: Φ1 จะสั่งยกเลิกการแยก Φ2 ได้ (A court will cancel to separate marriage property.)

In Ex. 2, EDU2 omits subject "ศาล" (court) and object 'สินสมรส' (marriage property). Therefore, word co-occurrence only is not enough to solve relationship between EDU1 and EDU2. This research use Absence rules to solve relationship between EDU1 and EDU2.

Implicit Marker Problem The absences of discourse marker in Thai language are often occurred. In Ex. 3, "แต่" (but) is a discourse marker which is omitted but relationship between EDU1 and EDU2 still have relationship similar to that EDU1 and EDU2 have the discourse marker.

Ex. 3. EDU1: ศาลได้มีคำสั่งให้แยกสินสมรส (A court has order to separate marriage property.)
EDU2: Φ1 ภริยาหรือสามีคัดค้าน (But a wife or a husband disagrees to separate marriage property.)

Therefore, only use cue phrase is not enough to solve relationship between EDUs such as EDU2.

C. *Issues in DR identification*

After RS trees are constructed correctly, DRs among EDUs need to be identified, and a number of issues need to be considered.

Adjacent Marker Problem Given three EDUs and two markers indicating different relationship, as shown in Ex. 1, there are two possible for the DR. First, relationship between EDU1 and EDU2 is "contrast" relation which is determined by using discourse marker "แต่" (but), next that between (EDU1, EDU2) and EDU3 is determined. On the other hand, the relationship between EDU2 and EDU3 is "condition" relation which is determined first by using discourse marker 'ถ้า' (if), next that between (EDU2, EDU3) and EDU1 is determined.

Implicit DR Marker Problem The absences of DR marker in Thai language are often occurred. In Ex. 3, "แต่" (but) is a "contrast" relation marker which is omitted but relationship between EDU1 and EDU2 still have "contrast" relation similar to that EDU1 and EDU2 have the DR marker. Therefore, only use DR marker is not enough to solve DR between EDUs such as EDU2.

DR Marker Ambiguity Problem one DR marker can have many semantics DR such as "เมื่อ" (when) can infer "condition" or "cause-result" relation, "แต่" (but) can infer "contrast" or "elaboration" relation in Ex. 4.

Identify applicable sponsor/s here. *(sponsors)*

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

Ex. 4. EDU1: ศาลได้มีคำสั่งให้แยกสินสมรส (A court has order to separate marriage property.)
EDU2: แต่กรียาหรือสามีคัดค้าน (But$_{\text{contrast relation}}$ a wife or a husband disagree,)
EDU1: ศาลได้มีคำสั่งให้แยกสินสมรส (A court has order to separate marriage property.)
EDU2: แต่ที่กรียาและสามีเห็นชอบ (But only$_{\text{elaboration relation}}$ a wife and a husband agree,)

III. Structures of Thai EDUs

A Thai EDU consists of infrastructure and adjunct constituents. There are twelve possible arrangements of Thai EDU [17], as shown in Table 1. The structure of an EDU "A teacher usually doesn't drink alcohol" is shown in Fig. 2.

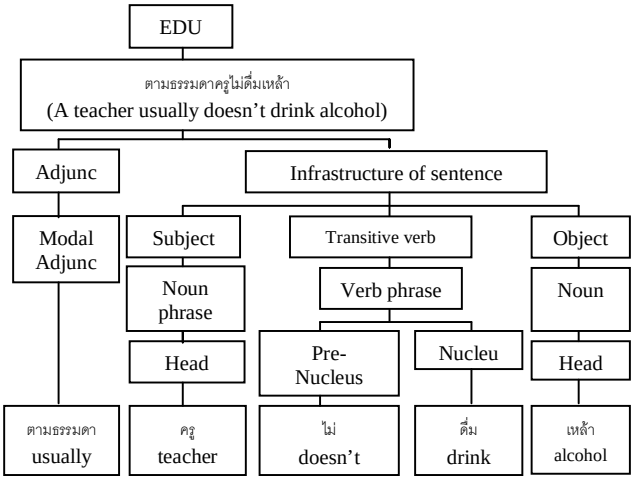


Fig. 2. The structure of the EDU "A teacher usually doesn't drink alcohol".

Table 1: The twelve patterns of Thai EDUs.

EDU	Examples	Rules
Vi	หิว (I'm hungry.)	NP_S-*Vi*-NP_S
S-Vi	ฝน-ตก (It's rain.)	
Vi-S	เจ็บไหม-คุณ(Are you pain?)	
Vt-O	หิว- น้ำ (I'm hungry.)	NP_O-NP_S-*Vt*-NP_O
S-Vt-O	รถ-ชน-เด็ก (The car hit the boy.)	
O-S-Vt	รูปนี้-ฉัน-ดูแล้วละจ๊ะ (I've already seen this photograph.)	
Vtt-O-I	ยังไม่ได้ให้-ยา-คนไข้ (I haven't given the patient the medicine.)	NP_S-*Vtt*-NP_O-NP_I
S-Vtt-O-I	ใคร-ให้-ลูกกวาด-หนู (Who gave you the sweet?)	
O-S-Vtt-I	ความลับ-ใครละ-จะกล้าถาม- คุณ (Who would dare to ask you the secret?)	NP_O-NP_S-*Vtt*-NP_I
I-S-Vtt-O	หนู-ป้า-จะให้-บ้านนี้ (Niece, I am going to give you this house.)	NP_I-NP_S-*Vtt*-NP_O
N	ป้า (Auntie)	NP_N-NP_N
N-N	นี่ปากกา-ใคร (Whose pen is this?)	

IV. EDU Segmentation

This section describes the EDU segmentation process proposed in this research. To reduce the segmentation ambiguity caused form modifiers, omissions of words and discourse markers, the first noun phrases and verb phrases which can be constituents of EDUs are determined, based on the possible structures. These phrases are then used to identify boundaries of EDUs.

A noun phrase (NP) is a noun or a pronoun and its expansion which may function as one of the four Thai EDU constituents, namely the subject (S), the object (O), the indirect object and the Nomen (N). The structure of a noun phrase consists of five constituents which are: head (H), intransitive modifier (Mi), adjunctive modifier (Ma), quantifier (Q), and determinative (D).

A verb phrase (VP) is a verb and its expansion which may function as one of the three Thai EDU constituents, namely intransitive verb (Vi), transitive verb (Vt) and double transitive verb (Vtt). The structure of a verb phrase consists of four constituents which are: nucleus (Nuc), pre-nuclear auxiliary (Aux1), post-nuclear auxiliary (Aux2), and modifier (M).

An Arrangement of NP and VP constituent [17] is shown in Table 2. There are twenty five possible arrangements of noun phrase and ten possible arrangements of verb phrases.

Table 2: The twenty five patterns of NP and ten patterns of VP.

Noun Phrase	Noun Phrase	Verb Phrase
H-Ma	H	Nuc
H-Mi-Ma	H-Mi	Nuc-Aux2
H-Q-Ma	H-Q	Nuc-M
H-Ma-Q	H-D	Nuc-Aux2-M
H-D-Ma	H-Mi-Q	Nuc-M-Aux2
H-Mi-Q-Ma	H-Q-Mi	Aux1-Nuc
H-Q-Mi-Ma	H-Mi-D	Aux1-Nuc-Aux2
H-Mi-D-Ma	H-Q-D	Aux1-Nuc-M
H-Q-D-Ma	H-D-Q	Aux1-Nuc-Aux2-M
H-D-Q-Ma	H-Mi-Q-D	Aux1-Nuc-M-Aux2
H-Mi-Q-D-Ma	H-Mi-D-Q	
H-Mi-D-Q-Ma	H-Q-Mi-D	
H-Q-Mi-D-Ma		

A. *Phrase Identification*

To perform the phrase identification, word segmentation and part of speech (POS) tagging are performed using SWATH [15] which extracts words and classifies them into 44 types such as common noun (NCMN), active verb (VACT), personal pronoun (PPRS), definite determiner (DDAC), unit classifier (CNIT) and negate (NEG). A hidden markov model (HMM) [14] employs these POS tag categories to determine phrases. The model assumes that at time step t the system is in a state PC (t) which has a probability of emitting a particular visible state POS tag (t), the transition probabilities a_{ij} among

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

hidden states and b_{jk} for the probability of the emission of a visible state:

$$\alpha_{\iota\varphi} = \pi(\Pi X_{\varphi}(\tau+1)/\Pi X_{\iota}(\tau)). \quad (1)$$

$$\beta_{\varphi\kappa} = \pi(\tau\alpha\gamma_{\kappa}(\tau)/\Pi X_{\varphi}(\tau)). \quad (2)$$

where PC (t) is Phrase Constituent at time step t, tag(t) is POS tag at time step t.

The transition probability for the hidden states where a particular sequence PC^T = {PC (1), PC (1), …, PC(T) } of T hidden states can be written as:

$$p(PC^T) = \prod_{t=1}^{T} p(PC(t)|PC(t-1)) \quad (3)$$

The hidden state and the corresponding visible state where the model generated the particular sequence of T visible POS tag state tag^T can write as:

$$p(tag^T|PC^T) = \prod_{t=1}^{T} p(tag(t)|PC(t)) \quad (4)$$

The probability that model produces a sequence tag^T of visible Pos tag states is

$$p(tag^T) = \arg\max_{PC_{1,n}} \prod_{t}^{T} p(tag(t)|PC(t))p(PC(t)|PC(t-1)) \quad (5)$$

The expression, p(PC(t)|PC(t-1)) is the probability of PC(t) given the previous PC(t-1), and p(tag(t)|PC(t)) is the probability of POS tag(t) given the Phrase Constituent(t).

This research use Baum-Welch [14] learning to determine model parameters, the transition probabilities a_{ij} and b_{jk}, from an ensemble of training samples.

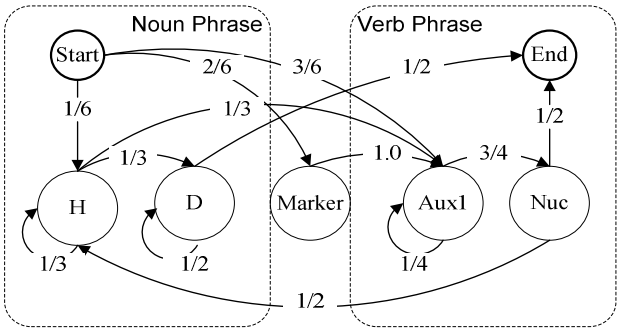


Fig. 3. A phrase model.

Given a sequence of visible state tag^T, the decoding is to find the most probable sequence of hidden states. This research uses Virterbi [14] to calculates p(tag^T) of visible POS states recursively because each term p(tag(t)|PC(t)) p(PC(t)|PC(t-1)) involves only tag(t), PC(t) and PC(t-1) by defining follows:

$$\delta_t(j) = \begin{cases} 0, t = 0 \; and \; j \neq initial \; state \\ 1, t = 0 \; and \; j = initial \; state \\ arg\; max_i \; \delta_{t-1}(i) a_{ij} b_{jkt} \quad otherwise, \end{cases} \quad (6)$$

where b_{jkt} represents the transition probability b_{jk} selected by the visible state emitted at time t. Thus the only nonzero contribution to the arg is for index k which matches the visible state tag(t).

	เพื่อน	จะขอ	ยืม	หนังสือ	เล่ม	นี้	
	Start	NCMN	XVMM	VACT	NCMN	CNIT	DDAC	END
Start	1	0	0	0	0	0	0	0
H	0	1/6*3/4	0	0	$8*10^{-3}$	0	0	0
D	0	0	0	0	0	$1*10^{-3}$	$3*10^{-4}$	0
Marker	0	2/6*0	0	0	0	0	0	0
Aux1	0	3/6*0	$3*10^{-2}$	0	0	0	0	0
Nuc	0	0	0	$2*10^{-2}$	0	0	0	0
End	0	0	0	0	0	0	0	$1*10^{-4}$
T =	0	1	2	3	4	5	6	7
Output	Start	H	Aux1	Nuc	H	D	D	End

Fig 4. The result of Viterbi tagging on the phrase model.

Figure 3 shows a phrase model of string "เพื่อนจะขอยืมหนังสือเล่มนี้ เพราะΦ_1ไม่ได้ซื้อΦ_2 ดังนั้นΦ3จึงต้องยืมหนังสือฉัน" (A friend's going to borrow this book. Because she (Φ_1) hasn't been able to buy it (Φ_2). Therefore she (Φ_3) must borrow it from me.) which POS tags of string is "เพื่อน (A Friend-NCMN) จะขอ (is going to-XVMM) ยืม (borrow-VACT) หนังสือ (book- NCMN) เล่ม (numerative-CNIT) นี้ (this-DDAC) เพราะ (Because-CONJ) เธอ (she(Φ_1)- PPRS) ไม่ (hasn't been-NEG) ได้ (able to-XVMM) ซื้อ (buy-VACT) มัน (it(Φ_2)) ดังนั้น(Therefore: CONJ) เธอ (she(Φ_3): PPRS) จึงต้อง (must: XVMM) ยืม (borrow: VACT) หนังสือ(book: NCMM) ฉัน (me: PPRS)". The hidden state of a phrase model consists of H(NCMN- book (2/4), friend (1/4); PPRS-me(1/4)), D (CNIT- numerative (1/2); DDAC- this(1/2)), Discourse-marker (CONJ- because(1/2), therefore (1/2)), Aux1 (XVMM- is going to(1/4), must(1/4), able to(1/4); NEG-hasn't been (1/4)) and Nuc (VACT- borrow (2/3), buy (1/3)). The hidden state of Thai EDU model consists of S (H- friend (1)), O (H- book (2/4); D- numerative, this (2/4)) I (H- me (1)), Discourse-marker (marker- because, therefore(1)) and Vt (Aux1- must, is going to, able to, hasn't been (4/7); Nuc-borrow, buy (3/7)).

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

B. *EDU Boundary Determination*

After we determine NP and VP, another HMM on EDU constituents (shown in Fig. 5.) is then created to determine the starting and the ending of boundaries of EDUs. This model can handle the subject and object absence problems, discussed earlier.

Fig. 5 shows an example of the EDU segmentation model of an EDU "เพื่อน-จะขอ-ยืม-หนังสือ-เล่ม-นี้" (A friend's going to borrow this book.) The EDU segmentation model can be expressed as:

$$p(tag^T) = \arg\max_{EDUC_{1,n}} \prod_{t}^{T} p(tag(t)|EDUC(t))p(EDUC(t)|EDUC(t-1)) \quad (7)$$

where EDUC(t) is EDU Constituent at time step t, tag(t) is Phrase tag at time step t.

The expression, p(EDUC(t)|EDUC(t-1)) is the probability of EDUC(t) is EDU constituent at time t given the previous EDUC(t-1) and p(tag(t)|EDUC(t)) is the probability of Phrase tag(t) given the EDU Constituent(t).

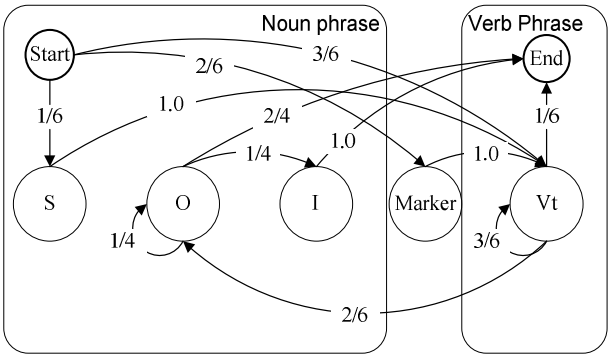


Fig. 5. A HMM of Thai EDU.

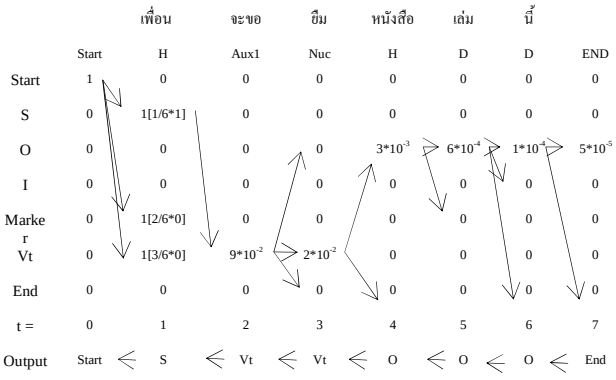


Fig.6. The result from Viterbi taged on EDU segmentation model.

C. *EDU Constituent Function Determination*

After we determine EDU constituent by using EDU segmentation model, rule base in Table 1 and Viterbi algorithm are then used to determine the grouping of function in EDUs. Evaluation of EDU constituent function determination is process which insert EDU constituent function tag for using in RS tree step. For example a string "เพื่อน-จะขอ-ยืม-หนังสือ-เล่ม-นี้" (A friend's going to borrow this book.), the result from Viterbi tagged on the EDU segmentation model is shown as: S, Vt, Vt, O, O, O

Rules in Table 1 are used to group constituents into appropriate functions.. For example, the rule based "NP_O-NP_S-Vt-NP_O" is used with label "S, Vt, Vt, O, O, O" because this label consists of Vt. Therefore, the result from Evaluation of EDU constituent function determination is shown as: NP_S – $(V, V)_t$ – $(NP, NP, NP)_O$

V. THAI RS TREE CONSTRUCTION

In this section, we describe our technique to construct an RS tree from a corpus based on rules derived from Thai linguistic characteristics. The rules are classified into three types which are Absence, Repetition and Addition [1, 3, 17, 18, 19]. Given a pair of EDUs, an author may write by using any or all of the three rules. The model calculates an EDU-EDU similarity matrices based on these sets of rules. Finally, a hierarchical clustering employs these matrices to build a rhetorical structure tree.

A. *Linguistic Rules for EDU Relations*

Absence rules In Thai language, it has been observed that frequently in writing some constituents of EDUs may be absent while the meanings are still the same. In the example below, the NP (object) "ขนม" (dessert) is absence from the anaphoric EDU according to rule, i.e., rule Φ (O, O).

Cataphoric EDU (VtO) : อยากจะทำขนมไหม (Would you like to make dessert?)
Anaphoric EDU (Vt) : อยากจะทำ (Yes, I like to make.)

Repetition rules It has been observed that frequently the anaphoric EDU is related to its cataphoric EDU by repetition of NP (subject, object), preposition phrase (PP) where it functions as a modifier of a nucleus or a verb phrase (VP). In the following example, two EDUs relate by a repetition of NP (object) "บ้าน" (House), i.e., rule я (O, O).

Cataphoric EDU (VttOI) : กำลังจะขายบ้านเขา (I'm going to sell him a house.)
Anaphoric EDU (VtO) : จะขายบ้านหลังไหน (Which house are you going to sell?)

Addition rules It has been observed that frequently the anaphoric EDU is related to its cataphoric EDU by addition of discourse marker, and possibly accompanied by Absence and

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

Repetition rules. In the example below Discourse Marker "เพราะ" (because) is added in front of the anaphoric EDU, i.e., rule Д (Marker, Before).

Cataphoric EDU (VttOI) : ฉันอยากจะยืมหนัง (I like to borrow films.)
Anaphoric EDU (VtO) : เพราะหาซื้อไม่ได้ (Because have not been able to buy it.)

Table 3 lists Repetition, Absence, and Addition rules. я (S, S) means that subject of the anaphoric EDU is repeated in the cataphoric EDU. Φ (S, S) means that subject is absence from the anaphoric EDU which subject appeared in the cataphoric EDU. Д (Marker, Before) means that Discourse Marker is added in front of the anaphoric EDU.

Table 3: Repetition, Absence, and Addition rules

Repetition (Я)	Absence (Φ)	Addition (Д)
я (S, S)	Φ (S, S)	Д (Marker, After)
я (O, S)	Φ (O, S)	Д (Marker, Before)
я (S, O)	Φ (S, O)	Д (Key Phrase, After)
я (O, O)	Φ (O, O)	Д (Key Phrase, Before)
я ((S, S), (S, O))	Φ ((S, S), (S, O))	
я ((S, O), (S, O))	Φ ((S, O), (S, O))	
я ((O, S), (S, O))	Φ ((O, S), (S, O))	
я ((O, O), (S, O))	Φ ((O, O), (S, O))	
я (S, Prep)	Φ (Only H, H)	
я (O, Prep)	Φ ((H, M), H)	
я (Prep, S)	Φ ((H, M), M)	
я (Prep, O)	Φ ((H, M), (H, M))	
я ((S, Prep), (S, O))	Φ (S, Prep)	
я ((O, Prep), (S, O))	Φ (O, Prep)	
я ((Prep, S), (S, O))	Φ (Prep, S)	
я ((Prep, O), (S, O))	Φ (Prep, O)	
я ((S, S), (S, Prep))	Φ ((S, Prep), (S, O))	
я ((S, O), (S, Prep))	Φ ((O, Prep), (S, O))	
я ((S, Prep), (S, Prep))	Φ ((Prep, S), (S, O))	
я ((O, S), (S, Prep))	Φ ((Prep, O), (S, O))	
я ((O, O), (S, Prep))	Φ ((S, S), (S, Prep))	
я ((O, Prep), (S, Prep))	Φ ((S, O), (S, Prep))	
я ((Prep, S), (S, Prep))	Φ((S, Prep), (S, Prep))	
я ((Prep, O), (S, Prep))	Φ ((O, S), (S, Prep))	
я ((Prep, Prep), (S, Prep))	Φ ((O, Prep), (S, Prep))	
я ((S, S), (O, Prep))	Φ ((Prep, S), (S, Prep))	
я ((S, O), (O, Prep))	Φ ((S, S), (O, Prep))	
я ((S, Prep), (O, Prep))	Φ ((O, S), (O, Prep))	
я ((O, S), (O, Prep))	Φ ((S, Prep), (O, Prep))	
я ((O, O), (O, Prep))	Φ ((S, O), (O, Prep))	
я ((O, Prep), (O, Prep))	Φ ((O, O), (O, Prep))	
я ((Prep, S), (O, Prep))	Φ ((O, Prep), (O, Prep))	
я((Prep, Prep), (O, Prep))	Φ((S, O, S), (S,O, Prep))	
я((S, O, Prep), (S, O, Prep))	Φ((S, S, O), (S,O, Prep))	
я (Only H, Only H)		
я (H, M)		
я (Only M, Only Nuc)		
я (Only M, Only M)		
я ((Nuc, M), (Nuc, M))		

B. *EDU Similarity Scoring*

A similarity score between two EDUs is calculated from contents of the EDUs, based on the three types of rules. A feature vector consists of Subject, Absence of Subject, Object, Absence of Object, Preposition, Absence of Preposition, Nucleus, Modifier Nucleus, Head, Absence of Head, Modifier Head, Absence of Modifier Head, Marker Before, and Marker After elements, each corresponding to a rule in the rule sets. The value of an element is dependent upon the type of rule, as follows:

Features based on absence rules:

A feature is generated from each absence rule with a value of:

$$If\ \ \Phi\ (C1, C2)\ is\ true\ then$$
$$value\ \ = 1 - \frac{|O1 - O2|}{Total\ number\ of\ sentences} \qquad (8)$$
$$Else,\ value\ \ = 0$$

where O1 is the order of cataphoric EDU, O2 is the order of anaphoric EDU, C1 is constituent of cataphoric EDU, C2 is constituent of anaphoric EDU.

Given the following example:

EDU1: ชาวบ้าน (/Subject) ประกอบ (/Nucleus) อุตสาหกรรมในครอบครัว (/Object) (The villager performs the family-industry.)
EDU2: และ (/Before) (/Subject, Absence of subject) หวงแหน (/Nucleus) สมบัติของชาติ (/Object) (and saves property of the nation.)
EDU3: อุตสาหกรรมในครอบครัว (/Subject) จึงเป็น (/Nucleus) สมบัติของชาติ (/Object) (Therefore, the family-industry is a property of the nation.)

In the example, the properties of EDU1 and EDU2 match with the rule "The anaphoric EDU2 is related to its cataphoric EDU1 by absence of subject", i.e., Φ (S, S), with the absence of subject "ชาวบ้าน" (Villager).

$$value_{EDU1,Subject} = 1 - \frac{|1 - 2|}{3} = value_{EDU2,Absence\ of\ Subject} \qquad (9)$$

Features based on repetition rules:

A feature is generated from each repetition rule with a value of:

$$If\ \ я\ (C1, C2)\ is\ true\ then$$
$$value = 1 - \frac{|O1 - O2|}{Total\ number\ of\ sentences} * \frac{Total\ \ of\ repeating\ words}{Total\ \ of\ words\ in\ sentences} \qquad (10)$$
$$Else,\ values = 0$$

where O1 is the order of cataphoric EDU, O2 is the order of

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

anaphoric EDU, C1 is constituent of cataphoric EDU, C2 is constituent of anaphoric EDU.

In the example, the properties of EDU1 and EDU3 match with the rule "The anaphoric EDU is related to its cataphoric EDU by repetition of subject", i.e., я (O, S), with the absence of object "อุตสาหกรรมในครอบครัว" (Family-Industry).

$$value_{EDU1,\ Object} = value_{EDU3,\ Subject} = (1 - \frac{|1-3|}{3}) * (\frac{1}{3} * \frac{1}{3}) \quad (11)$$

Features based on addition rules:

A feature is generated from each addition rule with a value of:

If Д (M, L) is true then
values = 1
Else, value = 0 (12)

Where M is Marker of EDU, L is link.

In the example, the properties of EDU1 and EDU2 match with the rule "The anaphoric EDU is related to its cataphoric EDU by addition of discourse marker "และ" (AND) ", i.e., Д (Marker, Before), link EDU2 with in front of EDU2 (EDU1). If addition rule is true, value will be assigned 1. But if addition is false, value will be assigned 0.

Similarity Calculation

Similarity between two EDUs (cataphoric and anaphoric) can be calculated as:

$$\Sigma\iota\mu\iota\lambda\alpha\rho\iota\tau\psi\ S_k(MR_{ij}) = \frac{max_k(MR_{ij}) - S_k(MR_{ij})}{max_k(MR_{ij}) - min_k(MR_{ij})} \quad (14)$$

where S_k (k \in {1, 2, 3}) is the normalized matrix whose values in matrix range from one (for the inter-EDU ranked best) down to zero (for the inter-EDU ranked worst), Matrix rule (MR) is a matrix from absence rule class, repetition rule class, and addition rule class, respectively; and Max's and Min's are taken over all occurring ranks for each rule [7]. The input score are normalized into range [0, 1] before computation takes standard combination algorithm [7]. S_i can be calculated separately for each type of rule, as follows:

Absence and repetition rules:

$$MP_{\iota\varphi} = |Magnitude_{EDU_i}^{Cataphoric} * Magnitude_{EDU_j}^{Anaphoric}| \quad (15)$$

where i and j are the order of EDU and |i-j| < MD

These rules consist of two parts (Cataphoric and Anaphoric). If two parts of absence and repetition rules are true, then absence and repetition rules are true. But if one part of absence and repetition rules is false, then absence and repetition rules are false.

Addition rule:

$$MP_{\iota\varphi} = |Magnitude_{EDU_i}^{Cataphoric} + Magnitude_{EDU_j}^{Anaphoric}| \quad (16)$$

Where i and j are the order of EDU and |i-j| < MD

This rule consists of two parts (Cataphoric and Anaphoric). If one part of addition rule are true, then addition rule are true.

The similarity above is used with a disjunctive hypothesis, where MD is the maximum distance of appropriate disjunctive hypothesis. Essentially, the disjunctive hypothesis enumerates relations of EDU_i over number of the Cartesian product {i, i + 1, …, i + MD + 1} X { j, j + 1, …, j + MD + 1}, i.e., all the pairs of EDUs that separated by an imaginary line drawn between EDU_i and EDU_j. The maximum distance 4 is the appropriate of disjunctive hypothesis.

VI. THAI RS TREE CONSTRUCTION

The methods used in this part are a hierarchical clustering. Because results from hierarchical clustering can be represent relation between EDUs in binary tree.

Each sample (an EDU in this case) begins in a cluster of its own and while there is more than one cluster left. The two closest clusters are combined into a new cluster and the distance between the newly formed cluster and each other cluster is calculated. The method of hierarchical clustering is shown in Table 4 and the result from hierarchical clustering is shown in Fig. 7.

Table 4. The hierarchical clustering method.

Clustering Method	Distance Between Clusters A and B
Single Linkage	The smallest distance between a sample in cluster A and a sample in cluster B
Un weighted Arithmetic Average	The average distance between a sample in cluster A and a sample in cluster B
Neighbour Joining	A sample in cluster A and a sample in cluster B are the nearest. Therefore, define them as neighbours.
Weighted Arithmetic Average	The weighted average distance between a sample in cluster A and a sample in cluster B.
Minimum Variance	The increase in the mean squared deviation that would occur if clusters A and B were fused

VII. THAI DR RECOGNITION

In this section, we describe our technique to recognize Thai DRs from a corpus based on rules derived from Thai linguistic characteristics. The rules are classified into three types which are Absence, Repetition and Addition [1, 3, 17, 18, 19]. A Decision tree (C5.0) employs these features to recognize a DR.

A. *Linguistic Rules for DR Recognition*

A feature score of DRs is calculated from contents of the EDUs, based on the three types of rules. A feature score consists of Cataphoric score (Subject, Object, Preposition, Nucleus, Marker Before and Marker After) and Anaphoric score (Subject, Absence of Subject, Object, Absence of

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

Object, Preposition, Absence of Preposition, Nucleus, Modifier Nucleus, Head, Absence of Head, Modifier Head, Absence of Modifier Head, Marker Before, and Marker After) elements, each corresponding to a rule in the rule sets. The value of an element is dependent upon the type of rule, as follows:

Features based on absence rules:

A feature is generated from each absence rule with a value of:

$$\begin{aligned} &\textit{If } Ø(C1,C2) \textit{ is true then} \\ &\quad Value_{Cataphoric} = 1 \\ &\quad Value_{Anaphoric} = \text{“Omit” } = 2 \\ &\textit{ELSE,} \\ &\quad Value_{Cataphoric} = 1 \\ &\quad Value_{Anaphoric} = -1 \end{aligned} \tag{17}$$

where C1 is constituent of cataphoric EDU, C2 is constituent of anaphoric EDU.

Given the following example:

EDU1: ชาวบ้าน (/Subject) ประกอบ (/Nucleus) อุตสาหกรรมในครอบครัว (/Object) (The villager performs the family-industry.)

EDU2: และ (/Before) (/Subject, Absence of subject) หวงแหน (/Nucleus) สมบัติของชาติ (/Object) (and saves property of the nation.)

EDU3: อุตสาหกรรมในครอบครัว (/Subject) จึงเป็น (/Nucleus) สมบัติของชาติ (/Object) (Therefore, the family-industry is a property of the nation.)

In the example, the properties of EDU1 and EDU2 match with the rule “The anaphoric EDU2 is related to its cataphoric EDU1 by absence of subject“, i.e., Φ (S, S), with the absence of subject “ชาวบ้าน” (Villager).

$$\begin{aligned} &value_{EDU1,\ Subject} = 1 \\ &value_{EDU2,\ Absence\ of\ Subject} = \text{"Omit"} = 2 \end{aligned} \tag{18}$$

Features based on repetition rules:

A feature is generated from each repetition rule with a value of:

$$\begin{aligned} &\textit{If } я(C1,C2) \textit{ is true then} \\ &\quad Value_{Cataphoric} = 1 \\ &\quad Value_{Anaphoric} = 1 \\ &\textit{Else,} \\ &\quad Value_{Cataphoric} = 1 \\ &\quad Value_{Anaphoric} = -1 \end{aligned} \tag{19}$$

where C1 is constituent of cataphoric EDU, C2 is constituent of anaphoric EDU.

In the example, the properties of EDU1 and EDU3 match with the rule “The anaphoric EDU is related to its cataphoric EDU by repetition of subject“, i.e., я (O, S), with the absence of object “อุตสาหกรรมในครอบครัว” (Family-Industry).

$$\begin{aligned} &value_{EDU1,\ Object} = 1 \\ &value_{EDU2,\ Subject} = 1 \end{aligned} \tag{20}$$

Features based on addition rules:

A feature is generated from each addition rule with a value of:

$$\begin{aligned} &\textit{If } Д(M, L) \textit{ is true then} \\ &\quad values_L = M \\ &\textit{Else, } value_L = -1 \end{aligned} \tag{21}$$

where M is Marker of EDU, L is link.

In the example, the properties of EDU1 and EDU2 match with the rule “The anaphoric EDU is related to its cataphoric EDU by addition of discourse marker “และ” (AND) “, i.e., Д (Marker, Before), link EDU2 with in front of EDU2 (EDU1). If addition rule is true, value will be assigned 1. But if addition is false, value will be assigned 0.

$$value_{EDU1,\ Before} = \text{”AND”} = value_{EDU2,\ Before} \tag{22}$$

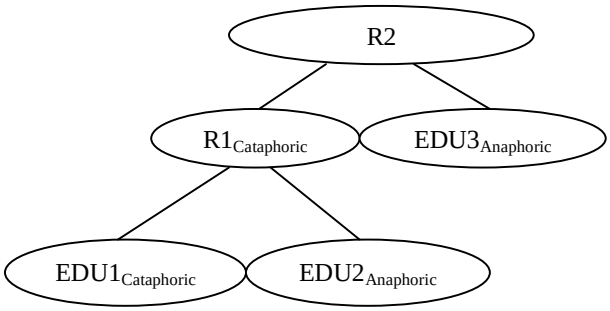


Fig. 7. R1 is DR of $EDU1_{Cataphoric}$ and $EDU2_{Anaphoric}$ by feature of $EDU1_{Cataphoric}$ consists of values {S, O, Prep, Nuc, $Marker_{before\ or\ after}$} and $EDU2_{Anaphoric}$ consists of values { S, O, Prep, Nuc, $Marker_{before\ or\ after}$ | $_{Repetition}$ ØS, ØO, ØPrep | $_{Absence}$ $Marker_{before\ or\ after}$ |$_{Addition}$ }, respectively and R2 is DR of $R1_{Cataphoric}$ and $EDU3_{Anaphoric}$ by R1 consists of values {S, O, Prep, Nuc, $Marker_{before\ or\ after}$ }$_{EDU1+EDU2}$ and $EDU3_{Anaphoric}$ consists of { S, O, Prep, Nuc, $Marker_{before\ or\ after}$ | $_{Repetition}$ ØS, ØO, ØPrep | $_{Absence}$ $Marker_{before\ or\ after}$ |$_{Addition}$ }, respectively.

VIII. Experimental Evaluation

A. *Evaluation of Thai EDU Segmentation*

In order to evaluate the effectiveness of the EDU segmentation process, a consensus of five linguists familiar with rhetorical structures of Thai texts is used to manually

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

segment EDUs of Thai family law which consists of 10,568 EDUs in total.

The EDU segmentation model is trained with 8,000 random EDUs, and the rest are used to measure performance.

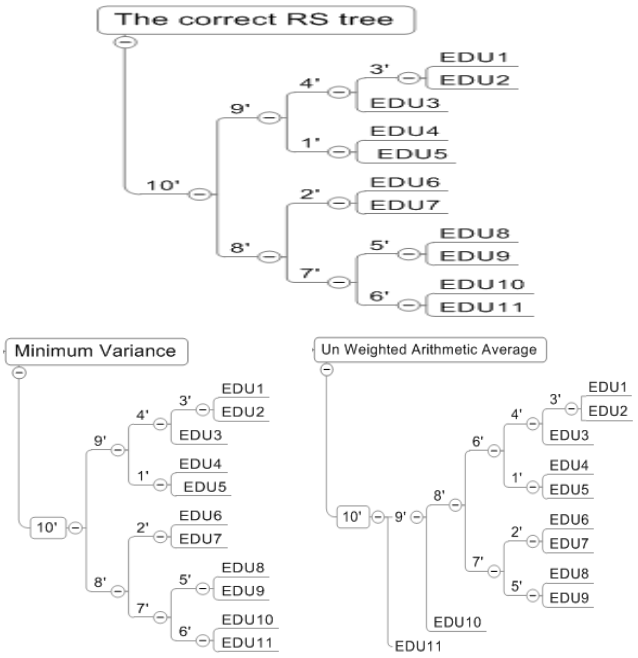


Fig. 8. Results from two hierarchical clustering algorithms

The training proceeds till the estimated transition probability changes no more than a predetermined value of 0.02 or the accuracy achieves 98%.

The performances of both phrase and EDU segmentations are evaluated using recall (Eq. 23) and precision (Eq. 24) measures, which are widely used to measure the performance.

$$Recall = \frac{\#\ correctlys\ Thai\ (Phrases;\ EDUs)\ identified\ by\ using\ model}{\#\ Thai\ (Phrases;\ EDUs)\ identified\ by\ analysts} \quad (23)$$

$$Precision = \frac{\#\ correctlys\ Thai\ (Phrases;\ EDUs)\ identified\ by\ using\ model}{Total\ number\ of\ Thai\ (Phrases;\ EDUs)\ identified\ by\ using\ HMM} \quad (24)$$

The results show that the proposed method achieves the recall values of 84.8% and 85.3%; and the precision values of 93.5% and 94.2% for phrase and EDU segmentations, respectively.

B. *Evaluation of EDU Constituent Function Determination*

In order to evaluate the effectiveness of grouping EDU constituents and determining their functions in an EDU using the proposed HMM model on NP and VP, the experiments employ three texts (Absence data with 84 EDUs, Repetition data with 117 EDUs and a subset of the Family law with 367 EDUs). The Absence data contains EDUs mostly following absence rules while the Repetition data contains those mostly following repetition rules. The analysts created training and testing data sets by manually building EDU constituents and inserting appropriate tags into each text.

The results of the Viterbi algorithm are labels of EDU constituents, then rules based on NP and VP are applied to group EDU constituents and insert function tags.

Table 5: Performance of determining EDU constituent functions

Rules	Absence data	Repetition data	Family law data
NP_S-Vi-NP_S	NP_S (100%)	NP_S (100%)	NP_S (100%)
NP_O-NP_S-Vt-NP_O	NP_S & NP_O (100%)	NP_S &NP_O (100%)	NP_S &NP_O (100%)
NPS-Vtt-NPO-NPI	NP_S &NP_O&NP_I (100%)	NP_S &NP_O&NP_I (100%)	NP_S &NP_O&NP_I (100%)
NP_O-NP_S-Vtt-NP_I NP_I-NP_S-Vtt-NP_O	NP_S (100%), NP_O&NP_I (91.37%)	NP_S (100%), NPO&NP_I (79.59%)	NP_S (100%), NP_O&NP_I (90.21%)
N-N	NP_N (100%)	NP_N (100%)	NP_N (100%)

Table 5 shows the results of grouping EDU constituents (Subject (S), Object (O), Indirect Object (I) and Nomen (N)) by using rules based on NPs assuming the position of verb phrases (Vi, Vt and Vtt) are known. In general, all rules except NP_O-NP_S-Vtt-NP_I and NP_I-NP_S-Vtt-NP_O perform well.

To resolve ambiguities with these two rules, a probability table of words in positions of NP_I and NP_O coming after position of Vtt ($P(Vtt|\ NP_I,\ NP_O)$) is used. The results of evaluating EDU constituents by using rules based on NP together with the probability table yield higher performance of NP_O and NP_I in absence data (92.24%), Repetition data (85.78%) and Family law (93.71%).

C. *Evaluation of Thai RS Tree Construction*

In order to evaluate the effectiveness of the Thai RS Tree construction process, linguists manually built the rhetorical structure trees of three texts which are Absence, Repetition and Family Law data sets, with a total of 568 EDUs. The performances of the rhetorical structure were evaluated by using recall and precision measures, i.e., (Eq. 25) and (Eq. 26) [6], respectively.

$$Recall = \frac{\#correctlys\ internal\ nodes\ identified\ by\ RST}{\#internal\ nodes\ identified\ by\ analysts} \quad (25)$$

$$Precision = \frac{\#correctlys\ internal\ nodes\ identified\ by\ RST}{Total\ number\ of\ internal\ nodes\ identified\ by\ RST} \quad (26)$$

In the Absence and Repetition data sets, though relations between EDUs follow mostly Absence rules and Repetition rules, respectively, many EDUs also follow other kinds of rules. For example,

Anaphoric EDU (SvtO) : บุรุษไปรษณีย์ (S) จะคัดเลือก (Vt) จดหมาย (ย O) (A Postman will select

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

letters)

Cataphoric EDU ((S)VtO) : และ (Д) (Φ S) จะรับส่ง (Vt) จดหมาย (я O) (And will receive and send letters)

Recall and precision are calculated with respect to the ability of an algorithm to construct an RS tree structure similar to that created by the linguists.

Table 6 shows example calculations of recall and precision of Thai RS Trees on a text created by the Minimum Variance and Unweighted Arithmetic Average algorithms in Fig. 8.

Table 7 shows the results of evaluating Thai RS Tree construction on the three data sets. The performance on the Family law which combines many kinds of rules in its content is 94.90% recall and 95.21% precision. The results also show that Unweighted Arithmetic Average clustering algorithm gives the best performance for Thai RS Tree construction.

A. *Evaluation of Thai DR Recognition*

In order to evaluate the effectiveness of the Thai DR recognition process, linguists manually built the DRs of Family Law data sets from RS tree, with a total of 1,248 EDUs. The DR model is trained with 828 random EDUs, and the rest are used to measure performance. The performances of the DR recognition were evaluated by using analysis module of Clementine version 12.0.

Table 8 shows the results of evaluating ten Thai DRs recognition on the data sets. The performance on the Family law which combines many kinds of rules in its content, found marker and not found marker is 82.50%, 85.09% and 81.28%, respectively.

Table 9 shows the results of comparison found marker and not found marker to recognize ten Thai DRs on the data sets. The performance on the Family law which found and not found marker in its content is 99.40% and 100%, respectively.

IX. CONCLUSIONS

Thai rhetorical structure tree (RST) construction is an important task for many textual analysis applications such as automatic text summarization and question-answering. This article proposes a novel two-step technique to construct Thai RS tree combining machine learning techniques with linguistic properties of the language.

First, phrases are determined and then are used to segment elementary discourse units (EDUs). The phrase segmentation model is a hidden markov model constructed from the possible arrangements of Thai phrases based on part-of-speech of words, and the EDU segmentation model is another hidden markov model constructed from the possible arrangements of Thai EDUs.

Table 6: Computing the performance of rhetorical structure tree (P=Precision; R=Recall).

The correct RS tree	Minimum Variance	Unweighted Arithmetic Average
/	3’	3’
/	4’	4’
/	1’	1’
/	9’	6’
/	2’	2’
/	5’	5’
/	6’	
/	7’	
/	8’	
		7’
		8’
		9’
		10’
	P =9/9	P=6/10
	R=9/9	R=6/9

As a side effect, functions of the EDU constituents can be determined by the EDU segmentation model together with linguistic rules grouping related constituents into a large unit. Experiments show the EDU segmentation effectiveness of 85.30% and 94.2% in recall and precision, respectively.

A hierarchical clustering algorithm with EDU similarity derived from semantic rules of the language is proposed to construct an RS tree. The technique is experimentally evaluated and the effectiveness is 94.90% and 95.21% in recall and precision, respectively.

Table 7: The results from RS tree experiment.

Data	N.	Clustering Method	Recall	Precision
Absence	84	Neighbour Joining	87.23	89.13
		Single Linkage	82.97	84.78
		Un weighted Arithmetic Average	87.23	89.13
		Minimum Variance	89.40	91.30
		Weighted Arithmetic Average	87.23	89.13
Repetition	117	Neighbour Joining	89.70	91.04
		Single Linkage	83.82	85.07
		Unweighted Arithmetic Average	89.70	91.04
		Minimum Variance	77.94	79.10
		Weighted Arithmetic Average	89.70	91.04
Family-	367	Neighbour Joining	85.98	86.26
Law		Single Linkage	64.01	64.21
		Unweighted Arithmetic Average	94.90	95.21
		Minimum Variance	63.37	63.57
		Weighted Arithmetic Average	90.44	90.73

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

Table 8: The results from DR recognition experiment.

Data	N.	Balance node (boost)	Correct
Family law	367	1244	82.81%
Not found Marker		715	81.28%
Found marker		529	85.09%

Table 9: Comparison DR of found and not found marker.

DR	Correct	Correct
	Not found Marker	found Marker
คล้อยตาม (consent)	93.10%	98.10%
ตัวอย่าง (example)	52.40%	54.00%
ลักษณะวิธี (characteristic)	69.40%	99.30%
สรุปความ (summary)	96.10%	-
เงื่อนไข (condition)	59.60%	85.30%
เลือกเอา (option)	97.70%	99.40%
เวลา (time)	62.50%	90.50%
เหตุผล (reason)	90.80%	91.20%
แจกแจง (explanation)	100.00%	-
แตกต่าง (contrast)	92.00%	98.90%

A decision tree (C5.0) algorithm with DR features derived from semantic rules of the language is proposed to recognize a DR of EDUs in RS tree. The technique is experimentally evaluated by setting boosting 10 number of trials, pruning severity 75, maximum record per child branch 2 and winnow attributes and the effectiveness is 82.81%.

REFERENCES

[1] D. Chamnirokasant, "Clauses in the Thai Language". Unpublished master's thesis, Chulalongkorn University, Thailand, 1969.

[2] D. K. Harman, editor. "The second Text Retrieval conference (TREC-2)", Gaithersburg, MD, USA, March 1994. U.S. Government Painting Office, Washington D.C, 1994.

[3] D. Mahatdhanasin, "A study of sentence groups in Thai essays". Unpublished master's thesis, Chulalongkorn University, Thailand, 1980.

[4] D. Marcu, "Build Up Rhetorical Structure Trees". In American Association for Artificial Intelligence, 1996.

[5] D. Marcu, "A decision-based approach to rhetorical parsing". In the 37th Annual Meeting of the Association for Computational Linguistics, ACL, Maryland, pp. 365-372 , 1999.

[6] D. Marcu, "The theory and Practice of Discourse Parsing and Summarization". The MIT Press, Cambridge, MA, 2000.

[7] E. A. Fox, and J. A. Shaw, "Combination of multiple search". In Harman [2], pages 243-249, 1997.

[8] J. Charoensuk, and A. Kawtrakul, "Thai Elementary Discourse Unit Segmentation by Discourse Segmentation Cues and Syntactic Information". In the Sixth Symposium on Natural Language Processing 2005 (SNLP 2005), Chiang Rai, Thailand, December 13-15, 2005.

[9] L. Alonso, and I. Castellón, Towards a delimitation of discursive segment for Natural Language Processing applications, International Workshop on Semantics, Pragmatics and Rhetorics, San Sebastián 22-24 November, 2001.

[10] L. Polanyi, "A formal model of the structure of discourse". Journal of Pragmatics, 12, 601-638, 1988.

[11] M. Wattanamethanont, T. Sukvaree, and A. Kultrakul, "Discourse relation recognition by using Naïve Bayesian classifier", The 9th National Computer Science and Engineering Conference (NCSEC 2005), University of the Thai Chamber of Commerce (UTCC), Bangkok, Thailand, October 27-28, 2005.

[12] R. Soricut, and D. Marcu, "Sentence Level Discourse Parsing using Syntactic and Lexical Information". In Proceedings of the 2003 Human Language Technology and North American Association for Computational Linguistics Conference (HLT/NAACL), May 27-June 1, Edmonton, Canada, 2003.

[13] R. Subba, , B. Di Eugenio, S. N. K. "Learning FOL rules based on rich verb semantic representations to automatically label rhetorical relations (EACL)", In Workshop on learning Structured Information in Natural Language Applications , 2006.

[14] S. Levinson, R. Rabiner, and M. Sondhi, "An introduction to the application of the theory of probabilistic function of a Markov proceeds to automatic speech recognition". Bell System Technical Journal, 62:1035-1074, 1983.

[15] T. Charoenporn, V. Sornlertlamvanich, H. Isahara, "Building A Large Thai Text Corpus---Part-Of-Speech Tagged Corpus: ORCHID---". In proceedings of the Natural Language Processing Pacific Rim Symposium, 1976.

[16] T. Sukvaree, J. Charoensuk, M. Wattanamethanont, and A. Kultrakul, "RST based Text Summarization with Ontology Driven in Agriculture Domain". Department of Computer Engineering, Kasetsart University, Bangkok, Thailand, 2004.

[17] V. Panupong, Inter-Sentence Relations in Modern Conversational Thai. The Siam Society, Bangkok, 1970.

[18] W. Aroonmanakun, "Referent Resolution for Zero Pronouns in Thai". Southeast Asian Linguistic Studies in Honour of Vichin Panupong. (Abramson, Arthur S., ed.). pp. 11-24. Chulalongkorn University Press, Bangkok. ISBN 974-636-995-4 , 1997.

[19] W. Aroonmanakun, "Zero Pronoun Resolution in Thai: A Centering Approach". In Burnham, Denis, Interdisciplinary Approaches to Language Processing: The International Conference on Human and Machine Processing on Human and Machine Processing of Language and Speech. NECTEC: Bangkok, 127-147, 2000.

[20] W. C. Mann, and S. A. Thompson, "Rhetorical structure theory. Toward a functional theory of text organization". Text, 8(3): 243-281, 1988.

AUTHORS PROFILE

Somnuk Sinthupoun is a teacher at the Department of Computer Science, Maejo University, Thailand. He holds a M.S. in Computer Science from National Institute of Development Administration (NIDA.), and a B. in Computer Science from Maejo University. His main research interests include artificial intelligence, information retrieval, data mining, and related areas.

Ohm Sornil is an Assistant Professor at the Department of Computer Science, National Institute of Development Administration, Thailand. He holds a Ph.D. in Computer Science from Virginia Polytechnic Institute and State University (Virginia Tech), an M.S. in Computer Science from Syracuse University, an M.B.A. in Finance from Mahidol University, and a B.Eng. in Electrical Engineering from Kasetsart University. His main research interests include computer and network security, artificial intelligence, information retrieval, data mining, and related areas.

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

Mobility Impact on Performance of Mobile Grids

A. S. Nandeppanavar[1], M. N. Birje[2], S. S. Manvi[3], Shridhar[4]
[1, 2]Department of ISE, Basaveshwar Engineering College, Bagalkot-587102, India
[3]Department of E&CE, Reva Institute of Technology & Management, Bangalore, India
[4]Department of E&CE, Basaveshwar Engineering College, Bagalkot-587102, India

.

***Abstract*— Wireless mobile grids are one of the emerging grid types, which help to pool the resources of several willing and cooperative mobile devices to resolve a computationally intensive task. The mobile grids exhibit stronger challenges like mobility management of devices, providing transparent access to grid resources, task management and handling of limited resources so that resources are shared efficiently. Task execution on these devices should not be affected by their mobility.**

The proposed work presents performance evaluation of wireless mobile grid using normal walk mobility model. The normal walk model represents daily motion of users and the direction of motion is mostly symmetric in a real life environment; thus it is effective in location updating of a mobile station and in turn helps task distribution among these available mobile stations. Some of the performance parameters such as Task Execution Time, task failure rate, communication overhead on Brokering Server and Monitoring Cost are discussed.

Keywords- Mobile Grids; Normal Walk Mobility Model; Mobility management; Collaborative Problem Solving; Task Management

I. INTRODUCTION

Mobile grid computing is one of the emerging grid types, with two underlying technologies: Mobile Computing and Grid Computing. Mobile grids focus on resource sharing among wireless mobile devices for distributed applications and are characterized by relatively high mobility and limited resources. Common resources shared are: computing power, data storage/network file system, communication and bandwidth, application software. These mobile grids exhibit stronger challenges like mobility management of devices, providing transparent access to grid resources, and handling of limited resources so that resources are shared efficiently. Therefore it is necessary to develop an environment model which can represent the issues mentioned above.

The work given in [1] focuses on wireless mobile grid for collaborative problem solving considering mobility related issues such as effect of mobility on performance of grid and network instability due to mobility. This mobile grid framework allows mobile devices to work collaboratively on computationally expensive tasks. Such a task is decomposed into smaller tasks and distributed across the other mobile devices willing to share their computational power with others.

Mobility models are important parameters for location updating of Mobile Stations (MS). We consider Normal Walk Mobility Model [2] to represent mobility pattern of users and to decide their location. The work also focuses on finding the direction of movement of MS and to predict instantaneously the Base Station Controller (BSC) with which handover occurs. Performance is analyzed based on parameters like task execution time, task failure rate, communication overhead on Brokering Server (BS) and monitoring cost.

The paper is organized as follows: Section II presents the related works on wireless mobile grid architectures, mobility models, mobility management and handovers in wireless mobile networks. Proposed work is described in section III. Section IV discusses about the results. And section V concludes the work.

II. RELATED WORKS

Ian Foster et al have defined grid [3] as "flexible, secure, coordinated resource sharing among dynamic collection of individuals, institutions, and resources what we refer to as virtual organizations". They have highlighted the need for grid technology in virtual organization.

T. Phan et al [4] have presented the challenge of integrating mobile devices with computational grid. The integration is provided through the use of an Interlocutor, which acts as proxy for cluster of Minions. There is no selection strategy to replace an interlocutor which has moved to another cell.

Kurkovsky et al [1], [5], and [6] have proposed an agent based approach to the design of wireless grid architecture to solve computationally expensive tasks. This architecture enables mobile devices within a wireless cell to form computational grid. It has several limitations, one of which being the inadequate consideration for the mobility of the mobile agents. Tasks are indiscriminately aborted by Subordinates and/or Initiators whenever these mobile agents move to neighboring cells.

P. Mudali et al [7] have proposed an extension to the architecture by Kurkovsky et al in [1]. They have proposed a multi-cell wireless computational grid, which is based on location area concept in GSM cellular networks. The proposed wireless computational gird is capable of greater device mobility tolerance than proposed in [1]. But still

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

there is need to introduce mobility management schemes in proposed architecture.

Ian F. Akyildiz [8] et al have proposed a simplified random walk model for hexagonal cell configuration where, probability states gives performance of the model. Further Guoliang Xue [9] and Md. Imdadul Islam and A.B.M. Siddique Hossain [10] have proposed improved models by reducing number of probability states to improve the performance.

Chiu-Ching Tuan et al have proposed a novel normal walk model for PCS networks with mesh cell configuration in [11] and compact normal walk model for hexagonal cell configuration in [2]. They represent the daily mobility behavior of an individual mobile station that moves from a cell to another, in PCS networks.

Jingyuan Zhang has described various location management schemes and mobility modeling method used for cellular networks in [12], and also provided the comparison of these methods.

M. N. Birje [13] et al have proposed a prediction based handover model for multiclass traffic in wireless mobile networks by using software agents, considering two cases: local handoff (between BSC's connected to same mobile switching center (MSC)), and global handoff (between BSC's connected to different MSC).

III. PROPOSED WORK

In this section we describe the architecture used in proposed work, mobility modeling, and location updating mechanisms.

A. Architecture

Figure 1 shows considered architecture. It includes two Virtual Organizations spanning over different Actual Organizations, which are monitored and controlled by a BSC. Its components are described briefly as follows:

Virtual Organization (VO): is the one which spans multiple actual organizations and transcends greater amount of geographical, organizational, and other type's related to intellectual property rights and national laws.

Actual Organization (AO): represents a single organization in a single place.

Base Station Controller (BSC): provides service to number of AOs through BTS. It has two servers BS and BSMS.

Base Transceiver Station (BTS): it supports the communication between BSC and mobile stations within AO. The BTS is fixed and is able to communicate with mobile stations using its radio transceiver.

Mobile Station (MS): is nothing but the mobile node which the user is holding. MS within an AO can play two roles: Initiator and Subordinator. Any device which is ready to take part in problem solving in grid environment is called as Subordinator. Initiator is the one which initiates or requests a task to be solved. Any subordinate may become an Initiator of distributed task if its user requests a large and/or computationally intensive task to be solved. In this case initiator is responsible for submitting such a distributed task to Brokering Server.

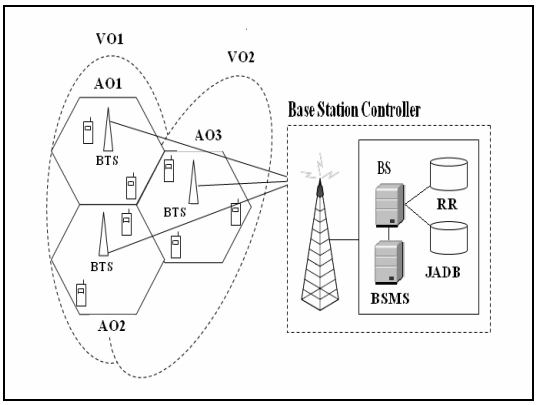


Figure 1. Architecture

Brokering Server (BS): it has the responsibility of task distribution and coordination in solving computationally intensive tasks. It knows the initial residing status of all devices in VO. Task distribution is based on available resources. BS has two data stores as below:

Resource Repository (RR): keeps track of available resources**.** Details of currently available resources of all mobile stations are stored in Resource Repository.

Job Allocation Database (JADB): keeps track of job distribution during task execution. Information about all subtasks allotted to different mobile stations and also results after performing operation are stored in this data base.

Base Station Monitoring Server (BSMS): It supports communication among mobile stations in wireless grid. It also keeps track of all the mobile stations available in the wireless grid. Each node entering or leaving the wireless grid should inform Base Station Monitoring Server. It maintains information about the mobile station's current location, which in turn helps to predict handoff in advance.

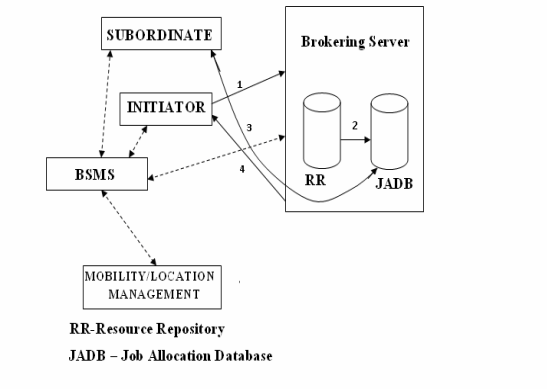


Figure 2. Interaction Diagram

Figure 2 explains the interaction among different components of the architecture. (1) Initiator sends the job to

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

be solved to the BS; BS looks for eligible nodes in RR, so that sub jobs can be distributed to them. (2) Details of sub job and nodes to which jobs are to be distributed are stored in JADB. (3) BS distributes sub jobs to eligible subordinates in the grid. Subordinates solve the sub jobs and return their partial results to the BS, which is stored in JADB (4) the initiator collects the partial results from the BS. BSMS supports the communication among all the nodes and servers. It is also responsible for location and mobility management of the mobile node.

B. *Mobility modeling and location updating*

Proposed architecture considers Normal Walk mobility model described in [11] for modeling movements of mobile stations within Actual Organizations. Normal Walk mobility model is a multi-scale, straight-oriented, mobility model. It represents the daily mobility patterns of a mobile station and the direction of motion is mostly symmetric in a real life environment. The mobile station moves in one step. Each move is based on previous move and is obtained by rotating previous move by an angle of θ in anticlockwise. This angle is called as moving angle, and helps to determine the next relative direction in which an MS (mobile station) moves across a cell in single step.

The drift angle θ in this model is a continuous random variable and its value helps to decide angle of movement of mobile station. The probability distribution of θ is assumed to approach normality rather than randomization with two parameters:

Mean (μ) with zero-degree.

Standard Deviation (σ) in the interval [5°, 90°]

Varying σ can redistribute the probabilities associated with θ and in turn helps to make the movement patterns more realistic to represent the user mobility. Thus any movement yielded from this model is function of σ, and is called a normal walk. The normal distribution of θ is represented as:

$$\theta \sim N(0°, \sigma^2) \qquad (1)$$

And probability density function of θ is defined by:

$$f(\theta)=\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{1}{2}(\frac{\theta}{\sigma})^2}, -270<\theta<270 \qquad (2)$$

This drift angle θ also helps to determine one of the six relative directions in which an MS handoffs from a hexagonal cell (in figure 3) to another in next step. The six directions are indexed by 'k' and given below:

- Making U-turn or turning back (B, k=0).
- Turning right (R, k=1).
- Moving front-right (Fr, k=2)
- Moving front or forward (F, k=3).
- Moving front-left (Fl, k=4)
- Turning left (L, k=5).

The direction k is not absolute for a mesh cell, but is relative to the inlet that an MS is currently visiting. The range of each direction k is confined to lie between two fixed angles. The confining angles are calculated as:

$$angF = \tan^{-1}\left(\frac{Ro}{4Ri}\right)$$

$$angFl = \tan^{-1}\left(\frac{Ro}{Ri}\right)$$

$$angL = \tan^{-1}(\infty) \qquad (3)$$

where Ri and Ro represent inner and outer radii of hexagonal cell respectively. Thus $angF = 16.1^\circ$, $angFl = 49.1^\circ$ and $angL = 90^\circ$.

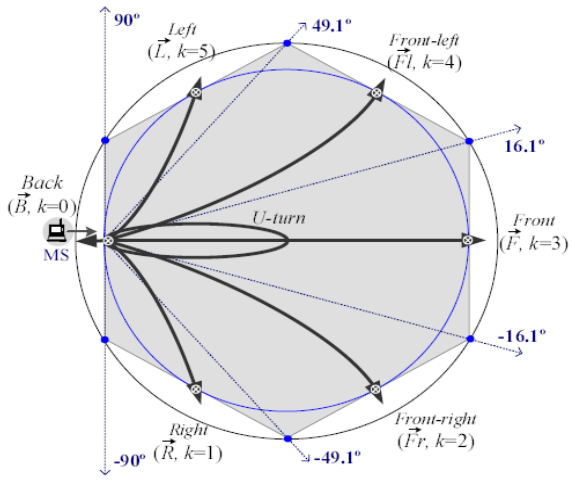


Figure 3. Hexagonal Cell

To find the probability of each of the six directions at next step, θ is standardized into Z (a continuous random variable), to find cumulative probabilities. Let Z= θ/σ such that Z has a standard normal distribution, Z ~ N (0, 1). The probability distribution function (pdf) and cumulative distribution function (cdf) are respectively given by

$$\varphi(z) = \frac{1}{\sqrt{2\pi}}\cdot e^{-z^2/2}$$

$$\phi(z) = \int_{-\infty}^{z}\varphi(\omega)d\omega \qquad (4)$$

Considering the equations (3), (4) the six probabilities associated with each direction k, with given σ are given by:

$$f = 2\cdot\phi(\frac{angF}{\sigma})$$

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

$$l = \phi(\frac{angL}{\sigma}) - \phi(\frac{angFl}{\sigma})$$

$$fl = \phi(\frac{angFl}{\sigma}) - \phi(\frac{angF}{\sigma})$$

$$b = 1 - 2\phi(\frac{angL}{\sigma})$$

$$fr = fl$$

$$r = l \qquad (5)$$

Movement of mobile node follows normal walk model, where BS knows the initial location of the device. Location updating of the device during movement follows the algorithm below:

1. Find the new location of the MS based on its current location and moving angle
2. If the MS moves out of VO, but within same AO, then information update is done to reflect its new location
3. If the MS moves out of VO, and also from local AO, then handover occurs

C. Job Distribution and Execution

BS of the proposed architecture is responsible for job handling, so that sub jobs of a given large job are distributed and partial results are collected back. BS has a data base called as JADB to keep track of job distribution among mobile devices within a wireless grid. It includes sub jobs distributed, identity of node to which jobs allotted and partial results. If a mobile device within wireless grid initiates a new job, the source code needed to run the corresponding job and all relevant parameters are submitted to the BS. BS generates sub jobs and assigns the sub jobs to available eligible subordinates, and waits to collect back the partial results.

Steps used to distribute and solve a computationally expensive job in the wireless grid environment can be described as below:

1. The user of a mobile device initiates a computationally intensive job
2. The Initiator creates sub jobs and transmits them to Brokering Service
3. The Brokering Service stores the received sub jobs in the JADB
4. The Brokering Service uses heuristics to find a subordinate in the RR for each sub job
5. The Brokering Service transmits sub jobs to the Subordinates
6. Each chosen Subordinate receives a sub job
7. Subordinates execute the sub jobs. During job execution
   a) if the subordinate moves out of VO, but within same AO, then the task assigned to that subordinate is continued to execute on it and the location of the subordinate is updated
   b) if the subordinate moves out of VO, and also from local AO, then task assigned to it is terminated. BS then redistributes it to another eligible subordinate and updates its JADB.
8. The Brokering Service receives partial results from subordinates that have finished their sub jobs and store them in the JADB
9. The initiator may chose to accept partial results or wait until all results are received from Brokering Service

The efficiency of the proposed work is measured using the average time taken for a task to be initiated, distributed, solved by the subordinates and returned back to the initiator. This average task execution time depends on different parameters like population of wireless grid and device mobility. Population of wireless grid specifies number of mobile nodes present in the wireless grid. Device mobility refers to the probability of a new device joining a wireless grid or a current device leaving the wireless grid per unit of time.

Some of the performance parameters evaluated are as follows:

Task execution time: It is the time taken for a task to be distributed by initiator, solved by the subordinates, and returned back to the initiator. It depends on different parameters like mobility factor (percentage of mobile nodes joining or leaving the wireless grid) of nodes within grid environment and grid population. To study the effect of grid population on execution time, the simulator is run considering different grid population. And to study the effect of mobility on execution time, the simulator is run under two scenarios: with and without mobility of nodes. In without mobility, nodes are considered to be at fixed location and tasks are distributed among them to find execution time. Where as in mobility, number of nodes is assumed fixed and mobility factor is varied.

Location Monitoring overhead: Location monitoring is keeping track of nodes status. BSMS has the responsibility of keeping track of nodes. It updates the database whenever a node enters or leaves the grid. Thus number of updates helps to determine the location monitoring overhead. The wireless grid simulator is run with different mobility factors and number of updates for each run is tabulated.

Communication overhead: location updating overhead leads to communication overhead because, location updating increases number of communications among BS

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

and BSMS. This in turn leads to increase in bandwidth utilization. Bandwidth utilization for different mobility factors is recorded to study its effect on communication.

IV. SIMULATION AND RESULTS

The proposed model has been simulated for various wireless grid scenarios. It considers *m* number of VOs controlled by a single BSC, with *c* number of AOs in each VO. The number of nodes in a VO varies between *n1* to *n2*. Each VO is allocated *v* Mbps bandwidth. Mobility factor in any VO is represented by *mf*.

The inputs considered for simulation are: $m = 2$, $c = 2$ for both VOs, $n1 = 30$ and $n2 = 90$. $v = 40$. mf = varies between 10% to 40%. Results are depicted using graphs as below:

Figure 4 shows the execution time calculated considering different grid size (grid population), where mobile nodes are stationary. The graph depicts that increase in grid size reduces the execution time i.e. increasing grid size in turn helps to provide more resources for executing the tasks and thus can be executed faster with more resources.

Figure 5 shows effect of mobility on execution time considering different mobility factors. The graph depicts that execution time gradually increases with an increase in mobility factor, because higher mobility factor leads to higher rate of task abortion and reallocation of task.

Figure 6 shows the effect of mobility factor on number of updates. The graph depicts that location update rate is more as mobility is more. Increasing mobility factor lead to increase in number of updates i.e. location update cost.

Figure 7 shows the effect of mobility factor on bandwidth utilization. The graph depicts that bandwidth utilization rate is more as mobility is more.

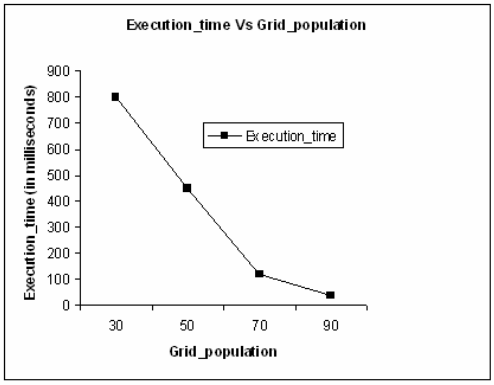


Figure 4. Execution time Vs. Grid_population

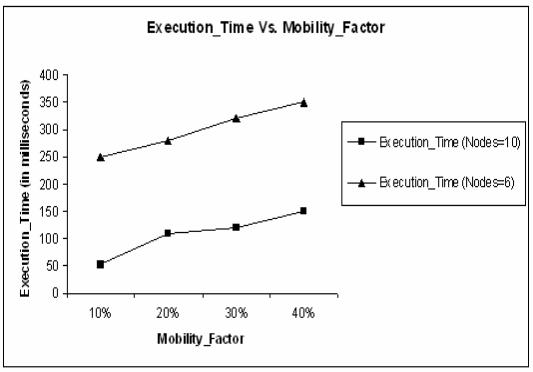


Figure 5. Execution time Vs. Mobility_factor

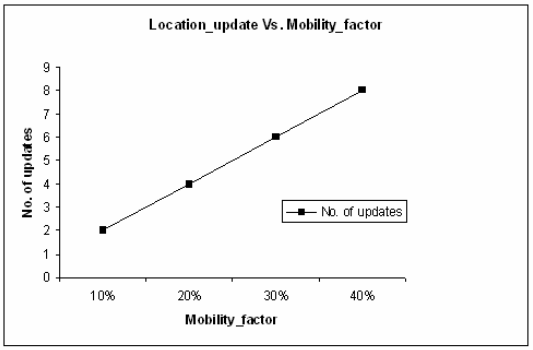


Figure 6. No. of updates Vs. Mobility_factor

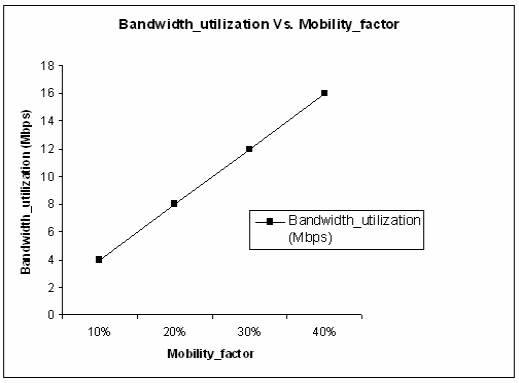


Figure 7. Bandwidth_utilization Vs. Mobility_factor

V. CONCLUSION

The proposed work presents implementation of normal walk mobility model in a wireless mobile grid environment for a Virtual Organization with two actual organizations. Normal walk mobility model helps to decide the next relative direction in which mobile station moves and in which it handoffs form one AO to another. This model in turn helps BSMS to keep track of number of mobile stations available in AOs; so that Brokering Server can distribute tasks among available nodes for tasks request sent to Brokering Server.

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

The proposed work considers Normal walk model because, it represents daily motion of users. The directions of motion are mostly symmetric in a real life environment, thus it is effective in location updating of a MS. The work also shows the effect of grid population and mobility factor on execution time and monitoring cost. It can be concluded that, high rate of resource availability decreases task execution time and higher mobility factor increases execution time. Finally increase in mobility factor increases monitoring overhead.

REFERENCES

[1] Stan Kurkovsky, Bhagyavati, Arris Ray, "A Collaborative Problem-Solving Framework for Mobile Devices", ACMSE April 2-3, 2004 Huntseville, Alabama, USA.

[2] Chiu-Ching Tuan and Chen-Chau Yang, " A Compact Normal Walk Model for PCS Networks" Mobile Computing and Communications Review, Volume 7, Number 4, 2004.

[3] Ian Foster, Carl Kesselman, and Steven Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", May 2001.

[4] T. Phan, L. Huang, C. Dulan, "Challenge: Integrating Mobile Wireless Devices into the Computational Grid", in Proceedings of the 8th International Conference on Mobile Computing and Networking, Atlanta, GA, September 2002.

[5] Stan Kurkovsky, Bhagyavati, "Modeling a Computational Grid of Mobile Devices as a Multi-Agent System", in Proceedings of the 2003 International Conference on Artificial Intelligence, Las Vegas, NV, June 2003.

[6] Stan Kurkovsky, Bhagyavati, Arris Ray, "Modeling a Grid-Based Problem Solving Environment for Mobile Devices", in Journal of Digital Information Management Vol.2 No.2, June 2004.

[7] P. Mudali, M.O. Adigun, and J.O. Emuoyibofarhe, "Minimizing the Negative Effects of Device Mobility in Cell-based Ad-hoc Wireless Computational Grids".

[8] I. F. Akyildiz, Y.-B. Lin, W.-R. Lai, and R.-J.Chen, "A new random walk model for PCS networks", IEEE J. Selected Areas Commun. 18, 7, pp. 1254-1260, July 2000.

[9] Guoliabg Xue, "An improved Random Walk Model for PCS Networks", IEEE Trans. Commun. 50, 8, pp. 1224-1226 August 2002.

[10] Md. Imdadul Islam and A.B.M. Siddique Hossain, "A proposed random walk model for mobile cellular network", December 2004.

[11] Chiu-Ching Tuan and Chen-Chau Yang, "A novel normal walk model for PCS networks with mesh cell configuration", 2005.

[12] Jingyuan Zhang, "Location Management in Cellular Networks", August 2004.

[13] M. N. Birje, S. S. Manvi, M. S. Kakkasageri, S. V. Saboji, "Prediction Based Handover for Multiclass Traffic in Wireless Mobile Networks: An Agent Based Approach" ICICS 2007.

[14] Ashish Agarwal, Douglas O. Norman, Amar Gupta, "Wireless Girds: Approaches, Architectures, and Technical Challenges", MIT Sloan School of Management, January 2004.

[15] WITSA (World Information Technology and Services Alliance), Background Paper on Grid Computing, September 2004.

[16] Buyurman Baykal, and Azgar B. Akan, "A Qos-Aware Handoff using RSVP in next generation Wireless Networks", Eighth IEEE Symposium on Computers and Communications, 2003.

[17] S. V. Saboji, S. S. Manvi, M. S. Kakkasageri, "Prediction based handover for multicast traffic in mobile networks", Proc. International Conference on Wireless Networks (ICWN-2007), The 2007 World Congress in Computer Science, Computer Engineering, Applied Computing, Las Vegas, Nevada, USA, pp. 293-298, June 25-28, 2007.

[18] Brijesh K R. Gupta, Mohan Lal, S. C. Sharma, "A Hybrid Scheme for Handover in ATM Based Personal Communication, vol. 3, no. 1, 2006.

[19] Cellular Networks, http//www.bridgeportnetworks.com/technology/wp_imshandover.pdf

[20] Jidong Wang and Lichun Bao, "Mobile Context Handoff in Distributed IEEE 802.11 Systems" IEEE Wireless Communications and Networking Conference (WCNC), pp. 225-230, 2004.

AUTHORS PROFILE

Anupama S. Nandeppanavar received B.E. in ISE from BLDEA's college of Engineering, Bijapur in 2005, and now she is PG student in Basaveshwar Engineering College, Bagalkot. Currently she is working as Lecturer Basaveshwar Engineering College, Bagalkot. Her interested area is computer networking.

Mahantesh N. Birje received B.E. in CSE from UBDT college of Engineering, Davangere in 1997, M.Tech. in CSE from Basaveshwar Engineering College, Bagalkot in 2005, and currently he is the research scholar at Visvesvaraya Techological University, Belgaum. He is working as Asst. Professor in the department of Information Science and Engineering, Basaveshwar Engineering College, Bagalkot Karnataka, INDIA. His area of interest include Grid computing, Multimedia Communications, and Agent technology. He has published 4 refereed journal papers, and 7 international conference papers.

Sunilkumar S. Manvi received M.E. degree in Electronics from the University of Visweshwariah College of Engineering, Bangalore, Ph.D degree in Electrical Communication Engineering, Indian Institute of Science, Bangalore, India. He is currently working as a Professor and Head of Department of Electronics & Communication Engineering, REVA Institute of Technology and Management, Bangalore, India. He is involved in research of Agent based applications in Multimedia Communications, Grid computing, Ad-hoc networks, E-commerce and Mobile computing. He has published 3 books, 3 book chapters, 25 refereed journal papers, and about 75 refereed conference papers. He has given many invited lectures and has conducted several workshops/seminars/conferences.

Shridhar received B.E. in E & CE from Gulbarga University in 1988, M.Tech. in digital electronics & advanced communications from KREC, Surathkal in 2000. He is currently working as Asst. Professor in the department of Electronics and Communications Engineering, Basaveshwar Engineering College, Bagalkot. His area of interest include signal processing and control systems.

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

Analysis of Birth weight using Singular Value Decomposition

D.Nagarajan
Department of Mathematics,
Salalah College of Technology,
Salalah,Sultanate of Oman.

V.Nagarajan
Department of Mathematics,
S.T.Hindu College,Nagercoil,
Kanyakumari,Tamil Nadu,India.

P.Sunitha
Department of Mathematics,
S.T.Hindu College,Nagercoil,
Kanyakumari,Tamil Nadu,India

V.Seethalekshmi
Department of Mathematics,
James College Technology,
Nagercoil,TamilNadu,India.

Abstract— **The researchers have drawn much attention about the birth weight of newborn babies in the last three decades. The birth weight is one of the vital roles in the baby's health. So many researchers such as [2],[1] and [4] analyzed the birth weight of babies. The aim of this paper is to analyze the birth weight and some other birth weight related variable, using singular value decomposition and multiple linear regression.**

Keywords- Birth weight; Haemoglobin concentration; Maternal Weight; Maternal height; Singular value decomposition.

I. INTRODUCTION

For a successful reproduction a good health throughout childhood, adolescence adult life and pregnancy is necessary. Special care has to be taken during pregnancy to get a healthy baby. Birth weight is an important determination of child health. It is a well known fact that the birth weight in influenced by the factors such as gestational period, maternal height, maternal weight, age, parity, haemoglobin concentration, rate of intrauterine growth, nutrition and many other socio-economic factors. Low birth weight neonate is defined as any neonate weighing less than 2500 grams at birth. It raises grave health risks for children which is a public health problem in most developing countries. Low birth weight stems primarily from poor maternal health and nutrition. Three factors have most impact poor maternal nutritional status before conception, short stature due mostly to under nutrition and infections during child hood and poor nutrition during pregnancy .Less than 17 years and greater than 34 years of age are at increased risk of low birth weight delivery. It can arise as a result of a baby being born too earlier (<37 weeks, also known as premature birth) or being born too small for gestational age (small as a result of intrauterine growth restriction). These types of babies have worst prognosis. Since low birth weight children are responsible for a very significant proportion of morbidity and mortality in childhood whereas child born with adequate birth weight are reported to do well even under adverse environment, researchers often use birth weight as a measure of morbidity risk.

II. DATA BASE

The data for the present study were collected from Chennai some private hospital during January 2008 to December 2008. The data for the selected variables were collected from the hospital case records.

III. DESCRIPTION OF MODEL

The singular value decomposition closely associated to the companion theory of diagonaling a symmetric matrix. Hark back that if A is a symmetric real n x n matrix there is an orthogonal matrix V and a diagonal D such that

$$A=VDV^{T}. \quad (1)$$

Here the columns of V are latent vectors for A and diagonal entries of D are eigen values of A for Singular Value Decomposition begin with m x n real matrix. There are orthogonal matrices U and V and a diagonal matrix S, such that

$$A = USV^{T}. \quad (2)$$

Here U is m x m and V is n x n, so that S is rectangular with the same dimensions as A .The matrix S can be formatted to be non negative and in order of decreasing order. The columns of U and V are called left and right singular vectors for A.[3]. Singular value decomposition is used in Latent Semantic Indexing (LSI) to determine the rank of the maternal variables and birth weight. Before scoring the maternal variables with Latent Semantic Indexing we need to onstruct a matrix with the maternal variables available as "A" .

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

TABLE I : MATERNAL VARIABLES BY BIRTH WEIGHT

Baby Weight	Maternal height (145-162)	Maternal Weight (50-70)	Age (21-35)	Blood Pressure (120/80)	Haemo globin (9-14)
< 2000	15	10	8	15	9
2000	78	55	58	35	60
2400	120	105	133	102	80
2800	206	240	230	180	150
3200	120	135	142	60	53
3600	8	21	15	25	19
>3600	3	7	4	10	3

The "Baby weight, Y" is taken as the dependent variable and other maternal variables are treated as independent variables X1, X2, X3, X4, X5, where

X1 : Maternal Height

X2 : Maternal Weight

X3 Age of mothers

X4 : Blood pressure

X5 : Haemoglobin concentration

Y : Baby weight

According to Singular Value Decomposition theory, an arbitrarily real rectangular matrix of order m x n can be decomposed into three matrices such that

$$A_{m\times n} = U_{m\times l}\, S_{l\times n}\, V^{T}_{n\times n}, \qquad (3)$$

where U and V are orthogonal matrices and S is a singular matrix with eigen values as its diagonal entries ,which are arranged in non - increasing order. The following analysis using MATLAB.

$$A = \begin{bmatrix} 15 & 10 & 8 & 15 & 9 \\ 78 & 55 & 58 & 35 & 60 \\ 120 & 105 & 133 & 102 & 80 \\ 206 & 240 & 230 & 180 & 150 \\ 120 & 135 & 142 & 60 & 53 \\ 8 & 21 & 15 & 25 & 19 \\ 3 & 7 & 4 & 10 & 3 \end{bmatrix}$$

$$U = \begin{bmatrix} -0.0424 & 0.1362 & -0.0520 & 0.0642 & 0.7905 & 0.5108 & -0.2948 \\ -0.2177 & 0.1121 & -0.8923 & -0.3417 & -0.0643 & 0.0783 & 0.1293 \\ -0.4183 & 0.2287 & -0.1740 & 0.8457 & -0.1576 & 0.0405 & 0.0284 \\ -0.7774 & 0.2715 & 0.3157 & -0.3517 & 0.1051 & -0.2749 & -0.0906 \\ -0.4088 & -0.8714 & 0.0160 & 0.0182 & -0.0222 & 0.2636 & 0.0547 \\ -0.0637 & 0.2748 & 0.2295 & -0.1896 & -0.5022 & 0.7609 & 0.0261 \\ -0.0201 & 0.0897 & 0.1348 & 0.0114 & 0.2869 & 0.0853 & 0.9400 \end{bmatrix}$$

$$S = \begin{bmatrix} 585.9583 & 0 & 0 & 0 & 0 \\ 0 & 49.1200 & 0 & 0 & 0 \\ 0 & 0 & 34.3363 & 0 & 0 \\ 0 & 0 & 0 & 21.2226 & 0 \\ 0 & 0 & 0 & 0 & 6.6886 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} -0.4737 & -0.1618 & -0.6429 & 0.1375 & 0.5632 \\ -0.5133 & -0.2822 & 0.4454 & -0.6588 & 0.1564 \\ -0.5230 & -0.3830 & 0.1032 & 0.5093 & -0.5565 \\ -0.3705 & 0.6792 & 0.4333 & 0.3596 & 0.2903 \\ -0.3182 & 0.5349 & -0.4358 & -0.3979 & -0.5142 \end{bmatrix}$$

From S matrix shows that its consists of five non zero singular values, confirming that A is a rank 5 matrix.

A. Dimensionality Reduction:

Computing U_K, S_K, V_K from U,S,V using MATLAB.

Let us take the economic dimension K =2 ,that is rank 2 approximation that means the first 2 columns of U and V and the first two rows and columns of S.

B. Decomposed matrices

Dimensionality reduction has been done my truncating the three matrices obtained form Singular Value Decomposition

$$U_K = \begin{bmatrix} -0.0424 & 0.1362 \\ -0.2177 & 0.1121 \\ -0.4183 & 0.2287 \\ -0.7774 & 0.2715 \\ -0.4088 & -0.8714 \\ -0.0637 & 0.2748 \\ -0.0201 & 0.0897 \end{bmatrix} \qquad S_K = \begin{bmatrix} 58595 & 0 \\ 0 & 49.12 \end{bmatrix}$$

$$V_K = \begin{bmatrix} -0.4737 & -0.1618 \\ -0.5133 & -0.2822 \\ -0.5230 & -0.3830 \\ -0.3705 & -0.6792 \\ -0.3182 & -0.5349 \end{bmatrix} \qquad S_K^{-1} = \begin{bmatrix} 0.0017 & 0 \\ 0 & 0.02035 \end{bmatrix}$$

It has been done by choosing the rank to be 2 ie, K = 2 is applied .Using

$$X_i = X_i^T U_K Sk^{-1}, \qquad (4)$$

and

$$Y = Y^T U_k Sk^{-1}, \qquad (5)$$

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

we get the new coordinate of vectors in this reduced space the new set of coordinate vectors are given below

$$Y = [-8.9121 \quad 7.4489]$$
$$X_1 = [-0.4709 \quad -0.1620]$$
$$X_2 = [-0.5107 \quad -0.2826]$$
$$X_3 = [-0.5205 \quad -0.3836]$$
$$X_4 = [-0.3681 \quad 0.6808]$$
$$X_5 = [-0.3164 \quad 0.5361]$$

Vector determination using cosine similarity values we rank results in decreasing order. Using

$$Sim(q,d) = \frac{q.d}{|q||d|} . \qquad (6)$$

Hence,

$$Sim(Y, X_1) = 0.516924$$
$$Sim(Y, X_2) = 0.360846$$
$$Sim(Y, X_3) = 0.237192$$
$$Sim(Y, X_4) = 0.929062$$
$$Sim(Y, X_5) = 0.94228$$

From the calculation $X_5 > X_4 > X_1 > X_2 > X_3$ .That means the value may be interpreted as the proportion of variability in Y that is explained by X_1, X_2, X_3, X_4, X_5 .It reveals that birth weight is closely related to Haemoglobin absorption. Haemoglobin is one of the vital roles in the birth weight of babies.

A comparison study is done between Singular Value Decomposition and multiple linear regression model to analyze the relationship between birth weight and maternal variable using the linear regression model of the form "Baby weight, Y" is taken as the dependent variable and the other variables are treated as independent variables.

TABLE II : REGRESSION COEFFICIENT

Predicator	Coefficient	t	VIF
Constant	3342.8	5.68	
X_1	-96.82	-1.09	283.827
X_2	9.05	0.22	75.032
X_3	68.39	0.78	309.698
X_4	-29.46	-0.58	59.953
X_5	45.02	0.56	106.884

From the above TABLE II it is observed that the value of R^2 is 0.738. It includes the maternal weight, height, age of mothers, blood pressure and Haemoglobin. It seems to 73.8% of the variation in baby weight is explained by the fitted model. The remaining 16.2 of variation can be explained by the factors other than these variables like socioeconomic factors.

Analyzing has been done for each maternal variable with the birth weight as shown below.

TABLE III: R^2 VALUES OF ALL INDEPENDENT VARIABLE

Dependent variable	Independent variable	R^2
Y	X_1	7.3
Y	X_2	1.5
Y	X_3	2.7
Y	X_4	3.0
Y	X_5	8.1

From the above Table III shows that the relation between Y and X1 is 7.3%, Y and X2 = 1.5, Y and X3 = 2.7, Y and X4 = 3.0 Y and X5 = 8.1

Hence also hemoglobin concentration is very close to birth weight .Hemoglobin is one of the vital role in the birth weight of babies.

TABLE IV: COMPARISON STUDY

Variable related to Y	Cosine Similarity	Multiple regression R^2
X_1	0.5169	7.3
X_2	0.3608	1.5
X_3	0.2371	2.7
X_4	0.9298	3
X_5	0.94228	8.1

Above table reveals that birth weight is very close to its haemoglobin concentration

IV. CONCLUSION

From this study, it is observed that, the birth weight mainly depends on the Haemoglobin concentration in both Sigular Value Decomposition analysis and multiple linear regression analysis. Hence the mother with high Haemoglobin concentration can avoid low birth weight. So the pregnant women should intake additional nutritional food to increase the Haemoglobin concentration and to avoid the health risk problems among the neonates. Some following reasons are the case of lowbirthweight, that is the mother has not obtained the appropriate nutrition, early marriage, late pregnancy at around 35 years, mother below 40 Kilograms, mother has anemia problems, small placenta, chronic placenta insufficiency can also lead to low birth weight , lack of Oxygen also leads low birth weight and due to mothers hypertension and malnutrition. Some action taken before birth to avoid low birth weight, which is regular checkup, appropriate nutrition intake, checks the haemoglobin status, take iron. The case should be identified and corrected in the

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

mother inability to gain weight should be deleted early to avoid problems. The following actions are taken after birth. In such circumstance the child needs extra care to maintain the temperature using like incubator, overhead warmer and kangaroo mother care.

REFERRENCE:

[1] Arbuckle T.E and Sherman G J,Comparison of the risk factors for pre terms delivery and intrauterine growth retardation ,Peachtree Perinat, Epidemol,(1989) Vol 3 ,pp115-129.

[2] Bhatia B.D and Tyagi N.K, Birth weight relationship with other Foelat Anthrometric parameter ,Indian pediatrics(1984),vol 21, pp833-838

[3] I.J Good, Some applications of the singular decomposition of matrix, Technometrics (1969).Vol 11,no 4,pp 823-831.

[4] Philip Steer u Kondaveeli c Bary-Kinsella, Relation between maternal haemoglobin concentration and birth weight in different ethic groups. Biomedical journal. (1995).310(1),pp 489-491.

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

A Simple method of designing dual loop controller for cold rolling mill

S.Umamaheswari, Dept. of EIE,
Mahendra Engg, College,
Namakkal(Dt), INDIA

V.Palanisamy
Principal/ Info Institute of Technology,
Coimbatore, INDIA

M.Chidambaram
Director/ NIT,
Trichy, INDIA

***Abstract*-The mathematical model (Interval Plant) of the web guide in rolling mill is controlled using PID controller. The given interval plant is approximated to first order plus time delay with integrator (FOPTDI) system. The dual loop control (DLC) method proposed by Jacob and chidambaram for design PID controllers is extended for FOPTD+I systems. The performance of the closed loop system is evaluated for both the original and the approximated model. The controllers are also tuned using Internal Model Control (IMC) and the performance is compared by simulation.**

***Index Terms* -Interval plant, Double loop control , Direct synthesis method, IMC(Internal Model Control)**

I. INTRODUCTION

There are many intermediate web guides in cold rolling mills process such as CRM (cold rolling mill), CGL(Continuous galvanizing line) and so on. The main functions of the web guide are to adjust the center line of the strip to the center line of the steel process, so they are called centre position control (CPL). Rapid process speed cause large deviation between the center position of the strip and the process line. So the difference between the center position of the strip and the process line should be compensated. In general, the centre position control (CPC) of the web is obtained by hydraulic driver and electrical controller. The model of the web guide system is obtained from Sang Min Kim et.al(2004)[1].

For the purpose of designing controllers, the dynamics of many processes can be described by first-order plus time delay with an integrator (FOPTDI) model.

Methods for tuning PID controllers for such models are based on stability analysis such as Internal model controller (Rivera et.al.(1986)[5]. The dual loop method is extended for designing PID controller for FOPTD+I systems. The inner loop is of PD controller and the outer loop is of PID controller designed using direct synthesis method proposed by shesagiri Roa[6]. PID controllers are also designed by IMC method. The performance of the controller is compared by simulation.

II. INTERVAL PLANT

The model of the web guide system by using geometrical relations of the guide ignoring the mass and stiffness of the web is given by Sang Min Kim et.al.(2004).

$$G(s) = \frac{a_3 s^3 + a_2 s^2 + a_1 s^1 + a_0}{s^5 + b_4 s^4 + b_3 s^3 + b_2 s^2 + b_1 s^1 + b_0}$$

Where the coefficients are of interval in nature, which is given by,

a_3 = [4.5 24] ; a_2 = [5 29] ; a_1 = [0.9 5] ; a_0 =[0.05 3]

b_5 = [0.05 0.3] ; b_4 = [4 9] ; b_3 =[4 9] ; b_2 = [0.7 2] ;

b_1 = [0.03 0.08]; b_0 = [0 0]

This is represented as model1, model2 and model3 by considering only the minimum, maximum and average values of the interval. The mode1, model2 and model3 of the plant are given by,

$$\text{Model1} = \frac{0.05 + 0.9s^1 + 5s^2 + 4.5s^3}{0.03s^1 + 0.7s^2 + 4s^3 + 4s^4 + s^5}$$

$$\text{Model2} = \frac{0.3 + 5s^1 + 29s^2 + 24s^3}{0.08s^1 + 2s^2 + 9s^3 + 9s^4 + s^5}$$

$$\text{Model3} = \frac{0.175 + 2.95s^1 + 17s^2 + 14.25s^3}{0.055s^1 + 1.35s^2 + 6.5s^3 + 6.5s^4 + s^5}$$

These models have been approximated to FOPTD+I system. Approximation is done by taking 1/s from the actual model separately, and applying step input, the time constant τ is calculated. Letting s to zero, the magnitude k is calculated.

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

Hence the FOPTD is obtained in which the 1/s term is again combined to get FOPTD+I model.
The approximated system is given by,

$$\text{RModel1} = \frac{1.667}{s(0.5584s+1)}e^{-0.3526s}$$

$$\text{RModel2} = \frac{3.75}{s(1.256s+1)}e^{-0.793s}$$

$$\text{RModel3} = \frac{3.1818}{s(1.0659s+1)}e^{-0.673s}$$

In the present paper dual loop system in which the inner loop is of PD controller and the outer loop is of PID for FOPTD+I system is designed. The performance of the control system is compared with that of the IMC method.

III. DUAL LOOP PID

The general transfer function of the process to be controlled is given by, $\frac{k_p}{s(\tau s+1)}e^{-Ls}$ (1)

where k_p is the process gain τ is the time constant and L is the process lag.

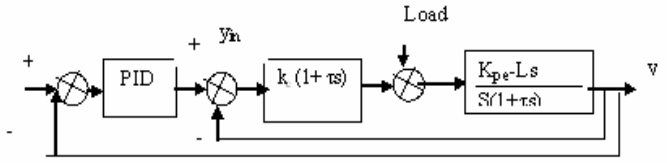


Figure 1. Block diagram of the dual loop control

The inner loop is of PI controller and the outer loop is of PID controller designed using direct synthesis method proposed by sheshagiri roa for unstable system. In this paper it has been extended for stable system. The closed loop Transfer function of the inner loop is given by, $y/y_{in} = \frac{k_c k_p e^{-Ls}}{s + k_c k_p e^{-Ls}}$, the delay is approximated to

$$e^{-Ls} = 1 - \left(L/2\right)s \Big/ 1 + \left(L/2\right)s$$

$$= \frac{k_c k_p\left(1-\left(L/2\right)s\right)}{s\left(1+\left(L/2\right)s\right)+k_c k_p\left(1-\left(L/2\right)s\right)}$$

$$= \frac{k_c k_p\left(1+\left(L/2\right)s\right)e^{-Ls}}{s\left(1+\left(L/2\right)s\right)+k_c k_p\left(1-\left(L/2\right)s\right)}$$

$$= \frac{k_c k_p\left(1+\left(L/2\right)s\right)e^{-Ls}}{L/2\,s^2+s\left(1-k_c k_p L/2\right)s+k_c k_p}$$

Closed loop transfer function of the inner loop is given by,

$$\frac{y}{y_{rin}} = \frac{k_c k_p\left(1+\left(L/2\right)s\right)e^{-Ls}}{\left(L/2\,s^2+s\left(1-k_c k_p L/2\right)s+k_c k_p\right)}$$

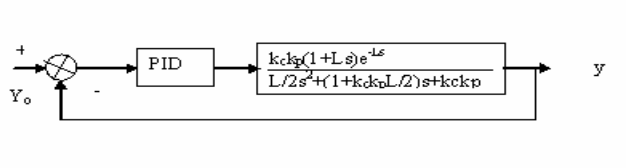


Figure 2. Block diagram of the simplified dual PID

after simplification the controller is obtained as,

$$G_c = k_c(1+\frac{1}{\tau_i s}+\tau_d s)\frac{(\alpha s+1)}{(\beta s+1)}$$

where

$$k_c = \frac{\eta_1}{k(3\lambda^2+1.5\theta\lambda+0.5\theta\eta_1-\eta_2)}$$

$$\tau_1 = \eta_1$$

$$\tau_d = \frac{\eta_2}{\eta_1}$$

$$\alpha = 0.5\theta, \beta = \frac{0.5\theta\lambda^3}{\tau(3\lambda^2+1.5\theta\lambda+0.5\theta\eta_1-\eta_2)}$$

in which and

$$\eta_2 = \frac{(0.5\theta-\tau)\lambda^3+(3\tau^2-1.5\theta\tau)\lambda^2+3\theta\tau^2\lambda+0.5\theta^2\tau^2}{\tau(0.5\theta+\tau)}$$

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

Tuning Parameter Selection

The tuning parameter λ should be selected in such a way that the resulting controller gains should be positive for positive values of k. Hence, as per the method given by Seshagirirao et. al. [6] to get positive values of controller gain(kc), the constraint to be followed is,

$$\eta < 3\lambda^2 + 1.5\theta\lambda - \eta_2 + 0.5\theta\eta_1$$

Also, λ should be selected in such a way that the resulting controller gives good robust control performances. The initial value of the tuning parameter can be taken as equal to half of the time delay of the process to get good control performances. If not then, the tuning parameter can be increased from this value till good nominal and robust control performances are achieved. For suitable value of λ and β , the controller designed on DS method gives good control performances. However for high value of β, the phase lag imposed by the term $(\beta s+1)$ in the controller is more and thus the designed controller with this value of β is not able to give robust control performances which results in low gain and phase margins of the open loop system than the required values (gain margin should be >1.7 and phase margin should be $>35^0$ for robust control of a process [6].

Based on many simulation studies, it is observed that taking '0.1 β' instead of β gives good compromise between nominal performances and robust control performances. Thus, in the present work, the value of β obtained is modified as '0.1 β' for simulation studies.

The performances of the closed loop system are evaluated by giving a unit step in the set point. Fig.1-3 shows the comparison of Rmodel 1, 2 & 3 with the controller tuned using double loop MRC method.

5. IMC METHOD

The IMC based PID controller settings for FOPTDI system is given by, (Dale, et. al. (2004) and Rivera et.al. (1986)[2&5] respectively.

$$k_c k = (2\tau_c + \tau + \theta)/(\tau_c + \theta)^2$$
$$\tau_I = (2\tau_c + \tau + \theta)$$
$$\tau_D = (2\tau_c + \theta)\tau/(2\tau_c + \tau + \theta)$$

The choice of design parameter τ_c is a key decision in both the direct synthesis (DS) and IMC design methods. In general increasing τ_c produces a more conservative controller because k_c decreases while τ_I increases. Several IMC guidelines for τ_c have been published for FOPTD with an integrator system. As per Rivera et al., (1986), τ_c/θ > 0.8 and τ_c > 0.1τ , $k_p \exp(-\tau_d s)/s(\tau s+1)$. By simulation it was found that τ_c =4 as per Rivera et. al., gives best result for all three models.

6. SIMULATION RESULTS

Let us consider the Rmodel2 plant with k_p =,3.75 τ = 1.256 and τ_d = 0.793. The PID controller settings by the dual loop method are k_c =0.452 τ_i = 0.96 and τ_D =0.366 The PID settings by IMC method are k_c =0.12 and τ_i =0.012 τ_D = 0.1319. Figure 1, 2 and 3 show the comparison of the servo response of all three models. The robustness of the closed loop is evaluated by using the maximum controller settings on the other two models of the system. Figures 4, 5 and 6 show the regulatory response of the systems. Comparison of ISE and IAE values of the Rmodel1, Rmode2 and Rmodel3 for both the servo problem and the regulator problems using the controller settings for the IMC and the dual loop method are shown in table I.

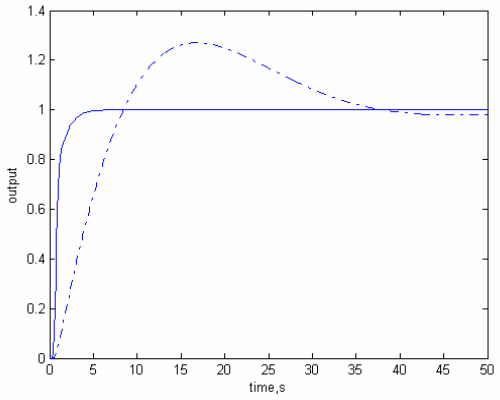


Fig.1 Servo response of Rmodel1
Solid -> DLC ; dash-dot –> IMC

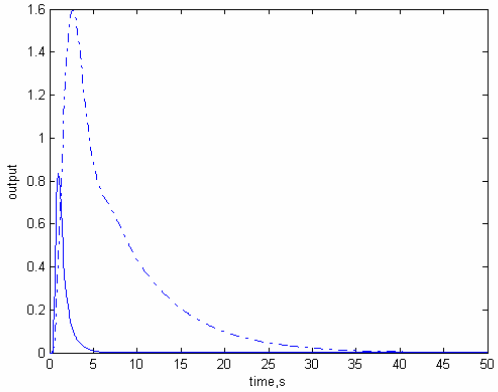


Fig.2 Regulator response of Rmodel1;
Legend as in Fig.1

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

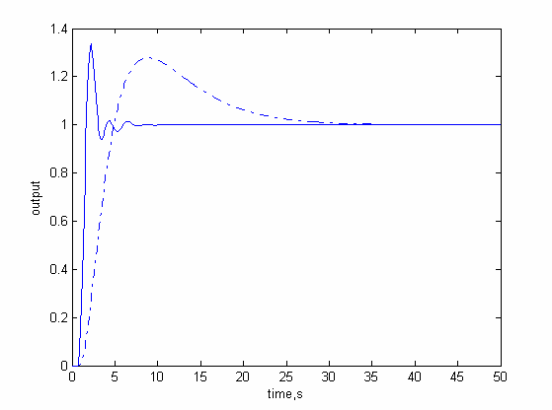


Fig.3 Servo response of Rmodel 2.
Legend as in fig.1

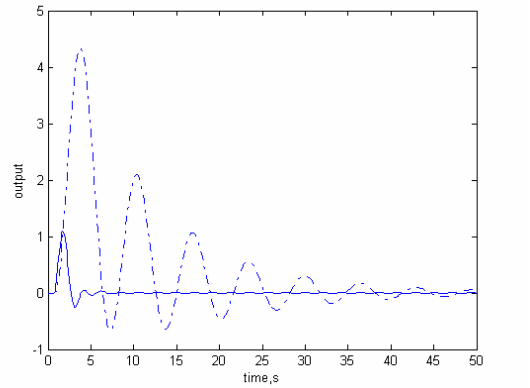


Fig.4 Regulator response of Rmodel 2
Legend as in fig.1

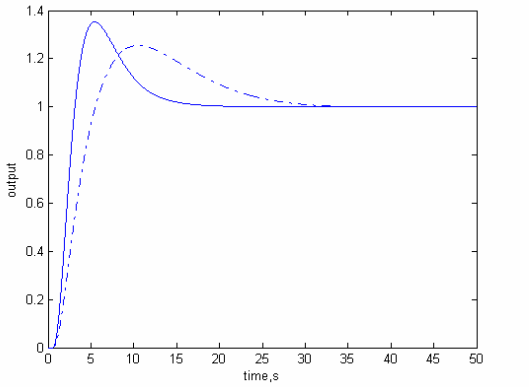


Fig.5 Servo response of Rmodel 3
Legend as in fig.1

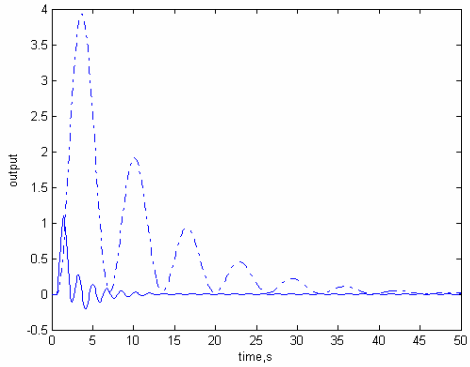


Fig.6 Regulator response of Rmodel 3
Legend as in fig.1

The servo and regulatory response of the dual loop control method gives a better response than that of IMC method. Table I and II gives the comparison of ISE and IAE values for both regulator and servo response. The performance of the controller designed by the dual loop method gives the best. Here the controller setting for the Rmodel2 is chosen and robustness is evaluated for other models.

TABLE I
ISE, IAE VALUE COMPARISON FOR REGULATOR RESPONSE

MODELS	METHOD	CONTROLLER SETTINGS			REGULATOR RESPONSE	
		K_C	K_C / τ_I	$K_{C^*} \tau_D$	ISE	IAE
MODEL 1						
	DUAL PID				3.79	498.6
	IMC				4.92	509.3
MODEL 2		0.452 0.12	0.96 0.012	0.366 0.131		
	DUAL PID				4.99	487.6
	IMC				6.78	500.9
MODEL 3						
	DUAL PID				4.98	501.2
	IMC				5.03	504.1

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

TABLE II
ISE, IAE VALUE COMPARISON FOR SERVO RESPONSE

MODELS	METHOD	CONTROLLER SETTINGS			SERVO RESPONSE	
		K_c	K_c / τ_I	$K_c * \tau_D$	ISE	IAE
MODEL 1						
	DUAL PID				0.116	9.62
	IMC				0.416	14.9
MODEL 2		0.452	0.96	0.366		
	DUAL PID	0.12	0.012	0.131	0.166	0.07
	IMC				0.33	120.92
MODEL 3						
	DUAL PID				0.14	0.17
	IMC				0.34	19.81

Also the same controller parameters are checked for their corresponding original model. The response shows that the dual loop method seems to be better compared to the other two models. The ISE and IAE values for regulator response are Reported in Table I. The dual loop method has got least ISE and IAE values.

7. CONCLUSION

The mathematical model of the web guide system is tuned using IMC and dual loop control method. The controller designed for FOPTD +Integrator system works well for even the corresponding original model. Also the robustness of the system is verified by simulation. The controller has been designed using dual loop control and IMC. Among this, robustness of the dual loop method is seems to be better when compared to IMC method. Referring to the Table I and II, the ISE and IAE values are comparatively less for dual loop method than that of the IMC method.

REFERENCES

[1] Sang Min Kim, Byoung joon Ahn, et.al., "The modeling and control of web guide process in cold rolling mill", IEEE Industrial Electronics Society, Korea, Nov2-6,(2004)

[2] E.Dale, Seborg, Jhomas F.Edgar, "Duncan A.Mellichamp, "PID controller design, tuning and troubleshooting", 2nd ed., pp. 304-309, John Wiley & sons(Asia) Pte.Ltd., (2004)

[3] M.Chidambaram and JacobE.F. "Control of unstable delay plus first order systems", Computer& Chemical Engineering, Vol.20, pp.579 -584, (1996)

[4] M.Chidambaram and JacobE.F. "Robust Control of an unstable bioreactor", process control & Quali., Vol.8, pp.167 - 175, (1996)

[5]Rivera,D.E.,Morari,M.&Skogestad.S,"IMC-PID controller design. Industrial Engineering and Chemical Process, Design and Development", 25, 252 (1986).

[6] Seshagirirao, A., V.S.R Rao, and M.Chidambaram, "Direct synthesis – based controller design for integrating processes with time delay," Journal of the Franklin Institute, 2008.

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

Detection of Microcalcification in Mammograms Using Wavelet Transform and Fuzzy Shell Clustering

T.Balakumaran
Department of Electronics and Communication Engineering
Velalar College of Engineering and Technology, Erode, TamilNadu, India.

Dr.ILA.Vennila
Department of Electrical and Electronics Engineering
PSG College of Technology, Coimbatore, TamilNadu, India

C.Gowri Shankar
Department of Electrical and Electronics Engineering
Velalar College of Engineering and Technology, Erode, TamilNadu, India.

***Abstract*— Microcalcifications in mammogram have been mainly targeted as a reliable earliest sign of breast cancer and their early detection is vital to improve its prognosis. Since their size is very small and may be easily overlooked by the examining radiologist, computer-based detection output can assist the radiologist to improve the diagnostic accuracy. In this paper, we have proposed an algorithm for detecting microcalcification in mammogram. The proposed microcalcification detection algorithm involves mammogram quality enhancement using multirresolution analysis based on the dyadic wavelet transform and microcalcification detection by fuzzy shell clustering. It may be possible to detect nodular components such as microcalcification accurately by introducing shape information. The effectiveness of the proposed algorithm for microcalcification detection is confirmed by experimental results.**

Keywords: Computer-aided diagnosis, dyadic wavelet transform, skewness and kurtosis, Fuzzy shell clustering.

I. INTRODUCTION

Breast cancer is the second leading cause of cancer deaths for women and is found as one in eight women in the United States. It is a disease in which cells in the tissues of the breast become abnormal and divide without order or control. These abnormal cells form too much tissue and become a tumor. According to WHO report, nearly two million women are diagnosed with breast cancer every year worldwide. The disease can be treated if discovered so early enough. The effective detection of breast cancer in earlier stage increases the survival rate. The appropriate method for early detection of pre cancerous symptoms is screening mammography, which has to be conducted as a regular test for women.

Calcification clusters are an early sign of breast Cancer. Microcalcifications are quite very small bits of calcium deposits present inside the soft breast tissue. It shows up in clusters or in patterns (like circles or lines) associated with extra cell activity in breast region.

Microcalcifications appear in mammogram image as small localized granular points with high brightness. It is not easy to detect by naked eye because of its miniaturized dimension. However about 10%-40% of Microcalcification clusters are missed by radiologists due to its small size and nonpalpable [1], [2]. To avoid these problems, a New CAD (computer Aided diagnosis) system has to be developed to improve the diagnostic rate. By incorporating the expert knowledge of radiologists, the CAD system can be made to provide a clear insight about the disease and saves the society from breast cancer.

Many researchers have proposed the algorithms for Microcalcification detection based on discrete wavelet transform, which is a powerful tool for analyzing the image hierarchically on the basis of scale. Some researchers have developed a CAD system using fuzzy clustering, artificial neural network and genetic algorithm. R. N. Strickland *et al.* [3], H.Yoshida *et al.* [4] used a wavelet transform to detect microcalcification and the fuzzy logic was tried by N.Pandey *et al.* [5]. Anne Strauss et al. [6] presented an identification scheme based on watershed Processing. Valverde *et al.* [7] used a deformable-based model for Microcalcification detection. Some of these studies detect approximately 70% to 80% of correct calcification. Objective of CAD system is to reduce the false positives and consistency of radiologists in image interpretation [8]. Naturally Microcalcifications are nodular in structure, other tissue such as mammary ducts blood vessels are linear in structure [9]. The Fuzzy shell clustering algorithm (FSC) is best in identifying circular objects present in an image [10]. In the proposed method, wavelet has been combined with fuzzy shell clustering (FSC) algorithm in order to mark the Microcalcification region

The rest of this paper is organized as follows. Section II presents microcalcification enhancement by using dyadic wavelet transform; the detection part using Fuzzy shell clustering in Section III, Results obtained on execution of algorithm are presented in section IV and Conclusion as last section.

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

II. MICROCALCIFICATION ENHANCEMENT

The fundamental operation needed to assist microcalcification detection in mammogram is image contrast enhancement. In many image-processing applications, the grayscale histogram equalization (GHE) is one of the simplest and effective techniques for contrast enhancement. Histogram equalization improves contrast of the microcalcification in mammogram image by reassigning the intensity values of pixels based on the image histogram. But Histogram equalization technique doesn't enhance the microcalcification region because it modifies the intensity of the image with some unpleasant visual artifacts.

Microcalcifications appear as subtle and bright spots, whose size varies from 0.3mm to 1mm in the mammogram image. It is not easy to enhance the microcalcification regions since surrounding dense breast tissue makes the abnormality areas almost invisible. Microcalcifications are high frequency in nature. So it can be extracted by using high pass filtering. But conventional enhancement technique like unsharp masking, homomorphic filters and high boost filtering tends to change the characteristics of microcalcification. To overcome these limitations microcalcification regions can be enhanced by dyadic wavelet transform without modifying characteristics of microcalcification.

Wavelet analysis permits the decomposition of image at different levels of resolution. In Fig. 1, the filter bank structure of the two-dimensional wavelet transform is shown from level j to level j+1, which generates four sub-images at level j+1. S_j be original image, the approximation sub-image S_{j+1} is obtained by applying the vertical low-pass filter followed by horizontal low-pass filter to S_j. The sub-image S_{j+1}^{LH} is obtained by applying the vertical low-pass filter followed by the horizontal high-pass filter. The sub-image S_{j+1}^{HL} is obtained by applying the vertical high-pass filter followed by horizontal lowpass filter. Finally, the response S_{j+1}^{HH} is obtained by applying the vertical and horizontal high-pass filters successively [9]. The downsampling by a factor 2 is introduced after each level of filtering. The same procedure is repeated for each level of approximation coefficients till S_{j+n} is achieved.

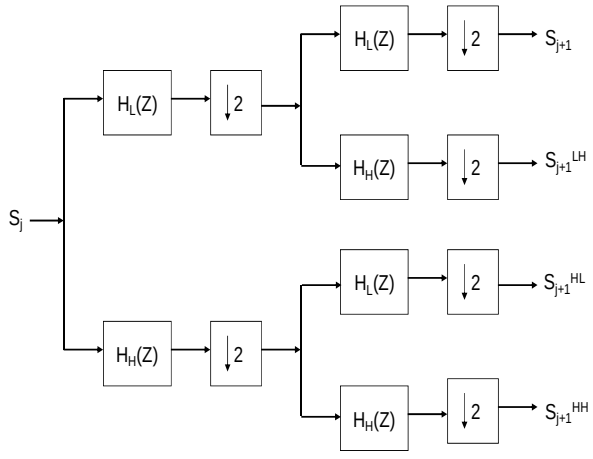


Fig. 1 Two dimensional dyadic wavelet structure

The digitized mammogram incorporated with a size of 1024 x 1024 pixel was taken from Digital Database for screening mammogram(DDSM). Mammogram image was decomposed upto 10 levels by applying dyadic wavelet transform with a decimation factor 2. The original 1024x1024 grayscale digital mammogram image was decomposed to 10 levels by applying Daubechies4 wavelet transform. Finally the lowest approximation image S_{10} is of single pixel width. Since microcalcification appears as high frequency behavior in mammogram, the enhancement is achieved by setting the value of S_{10} as zero.

The detail coefficients are enhanced by as per the expression number (1)

$$S_j^D(x,y) = \begin{cases} S_j^D(x,y) & \text{if } |S_j^D(x,y)| < T_j \\ g * S_j^D(x,y) & \text{if } |S_j^D(x,y)| \geq T_j \end{cases} \quad (1)$$

where, x and y are spatial coordinates, D represents all horizontal, vertical and diagonal subbands. T_j be a non negative threshold obtained by taking standard deviation of respective subimage. The best visual quality of microcalcification is obtained while the gain(g) is set as 1.2. The reconstruction of weighted higher frequency subbands provides better visibility of microcalcification region than the other breast regions.

III. MICROCALCIFICATION DETECTION

The CAD system was developed to assist the radiologist in detecting the suspicious areas on the basis of certain feature extracted from the images. After enhancing the microcalcification regions, next step in the journey of climbing the aim is the extraction of microcalcification through Soft computing tools. Microcalcification appears in mammogram as nodular points with higher brightness, localized or broadly diffused along the breast tissue, whereas normal tissues such as blood vessels are linear in structure. So detecting the nodular structure in image is a key in detecting the microcalcification.

A. Region of Interest (ROI) Identification

The first stage of microcalcification detection is ROI identification. The enhanced mammogram image is decomposed by undecimated wavelet transform (filter bank implementation without downsampling). The resulting horizontal detailed image or vertical detailed image is used to identify the region encircling the microcalcification clusters. Third and fourth order statistical parameters, skewness and kurtosis [11], are used to find the regions of microcalcification clusters.

An estimate of the skewness is given by

$$S_k = \frac{\sum_{i=1}^{N}(x_i - \tilde{m})^3}{(N-1)\sigma^3} \quad (2)$$

and the statistical parameter kurtosis holds the expression

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

$$K_u = \frac{\sum_{i=1}^{N} (x_i - \tilde{m})^4}{(N-1)\sigma^4} - 3 \quad (3)$$

where x_i is the input data over N observations, $\tilde{m}$ is the ensemble average of x_i and σ with its standard deviation. The third and fourth order statistical estimates were calculated for every overlapping 32x32 square regions of horizontal bandpass image or vertical bandpass image. The area having skewness value greater than 0.2 and kurtosis value greater than 4 is marked as a region of interest (ROI).

B. Fuzzy Shell clustering

Extraction of features is the key process in the development of CAD system. Fuzzy shell clustering (FCS) algorithm is used to perceive nodular structure from the ROI. It extracts the clusters with spherical symmetry of a given data set $X=\{x_1,x_2,.....x_n\}$ with the prototype cluster centers $V=\{v_1,v_2,.....v_c\}$. Here c denotes the number of clusters formed during processing. Each data point x_k has a degree of membership u_{ik} to the i^{th} cluster. The data point x_k is assigned to a cluster if the below given weighted objective function is minimum.

$$J = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^{m} Dik^2 + \sum_{i=1}^{c} W_i \sum_{k=1}^{n} (1-uik)^m \quad (4)$$

Where D_{ik} is distance from data point x_k to cluster center v_i and W_i be the 3-db width of the i^{th} cluster. The weighting exponent $m\in(1,\infty)$ is the fuzziness controlling parameter with m as a real number greater than 1, it was chosen as 2 for better segmentation. Microcalcification appears as nodular (circular) in structure, the FCS algorithm identifies nodules described by their center v_i and radius r_i. The geometric distance function of clustering is

$$D_{ik} = \left| \left\| x_k - v_i \right\| - r_i \right| \quad (5)$$

The necessary condition on membership leads to following equation used to update u_{ik} for the minimization of objective function

$$U_{ik} = \frac{1}{\sum_{i=1}^{c} \left(D_{ik} / Wi \right)^{\frac{2}{m-1}}} \quad (6)$$

The following equation is used to speed up the algorithm

$$\left\| u^{j+1} - u^{j} \right\| < \varepsilon \quad (7)$$

ε is minimum threshold value, which is used to minimize the number of iteration.

C. Nodular extraction

After identifying ROI, the next step is to detect edges of the enhanced image. Edges are detected by applying first order derivatives. First order derivative is implemented by gradient operator (G_x&G_y) . The edges are detected by computing the gradient of each pixel in the enhanced image in x direction and y direction. The gradient of f(x,y) is

$$\nabla f = \left| G_x \right| + \left| G_y \right| \quad (8)$$

By applying first order derivative, Enhanced edges were detected. These enhanced edge pixels are connected in sets to form group. The Groups containing less than 5 pixels are discarded. Fuzzy Shell clustering algorithm was applied to only edge point, which belongs to ROI. FCS was applied several times with c= 1 and W_i = 9. As a result, different circles with different radius were obtained. Final goal is to extract only circular (nodular) region, whose radius is equal to microcalcification radius.

After obtaining different circular regions, microcalcification detection is performed to distinguish between valid microcalcification region and invalid one. The two validity measurement parameters are Cluster density and Relative Shell Thickness [12]. These two parameters are used to identify microcalcification regions.

$$C_{di} = \frac{\sum_{k=1}^{n} u_{ik}}{2\pi r_i} \quad (9)$$

Numerator denotes sum of membership function on most characteristic points ($|u_{ik}| > 1/2$) and denominator denotes area of the circular region.

Relative shell Thickness is defined by

$$RST_i = \frac{\sum_{k=1}^{n} u_{ik}^{m} D_{ik}^{2}}{r_i \sum_{k=1}^{n} u_{ik}^{m}} \quad (10)$$

Where r_i is radius of circular region. Correct microcalcification was detected according to the rule

If (C_{di} >1.15 && RST_i < 0.2)
    Microcalcification nodular
End

The above thresholds were found by different experiments. Microcalcification was successfully detected by our proposed method.

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

IV. RESULTS

To test the proposed method, experiments were performed on the set of mammogram image with different size and features which were obtained from DDSM database. In DDSM database, the size of each pixel is 0.0435mm and gray level depth is 12 bits.

Fig.2(a) shows a low contrast mammogram image of size 1024x1024. Fig.2(b) shows enhanced image where one can clearly visible microcalcification than the original image. Fig.2(c) shows Regions of Interest (ROI) was calculated by high order correlation parameters such as Skewness and kurtosis. Fig.2(d) shows edge map of enhanced image by applying first order derivatives. Fig.2(e) shows edge map after removal of small unwanted groups and Fig.2(f) shows the microcalcification region after applying FCS. Here FCS was applied several times to each edge point which is belonging to ROI.

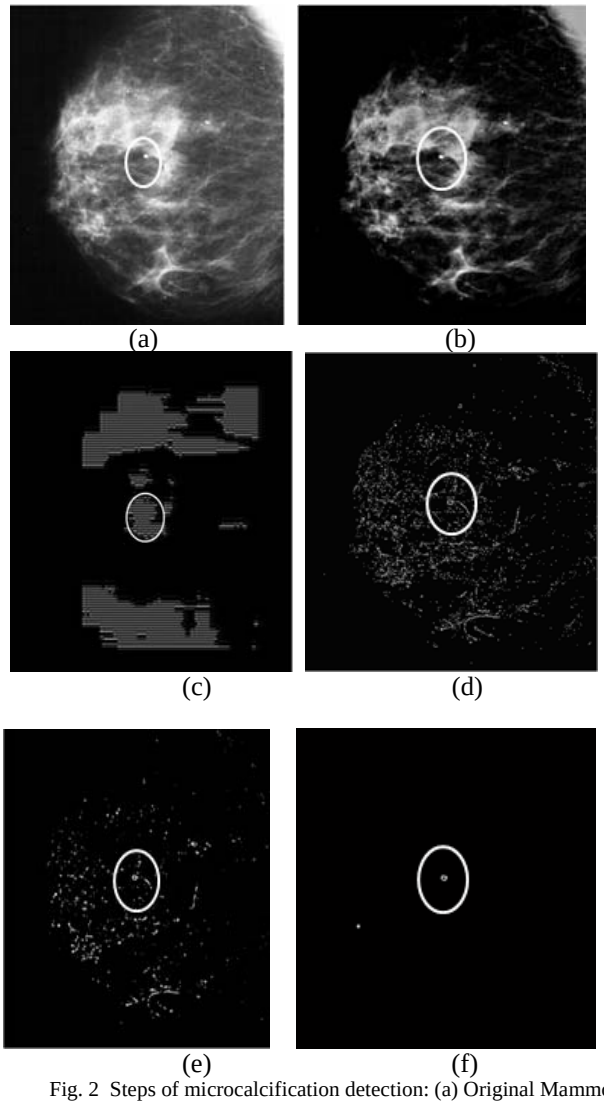


Fig. 2 Steps of microcalcification detection: (a) Original Mammogram Image (b) Enhanced image by wavelet transform (c) Region of Interest (ROI)identified by using Skewness & kurtosis (d) Edge map of enhanced image (e) Edge map after removal of unwanted groups (f) Microcalcification Detection by FCS algorithm

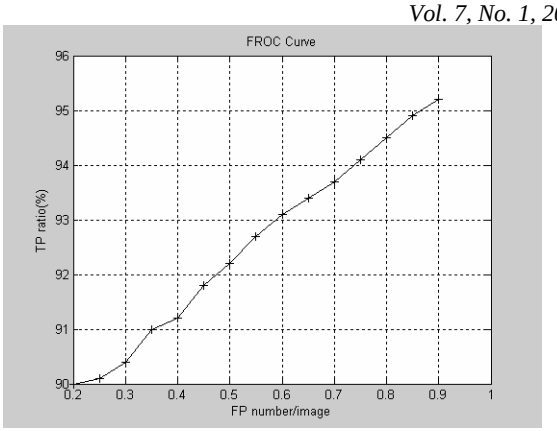


Fig. 3 FROC Curve of Microcalcification Detection

Free-response receiver operating characteristic (FROC) curve is used to evaluate the performance of microcalcification detection [13]. An FROC plot is achieved by calculating of the true-positive detection ratio (TP) with average number of false positives (FPs) per image. True positive detection ratio here refers that how many true abnormalities (microcalcifications) are correctly detected by computerized scheme and false positive per image refers to how many true abnormalities are missed. We have tested 112 images for FROC analysis. Our proposed method was achieved 95.23% TP ratio with 0.9 FP number/image.

V. CONCLUSION

In this paper, an algorithm for image enhancement based on wavelet transform and microcalcification detection using a Fuzzy shell clustering were proposed. Original Mammogram image is decomposed by applying dyadic wavelet transform. Enhancement is performed by setting lowest frequency subband of decomposition to zero. Then image is reconstructed from the weighted detailed subbands, the visibility of the resultant's image is greatly improved. Experimental results were confirmed that microcalcification can be accurately detected by fuzzy shell clustering algorithm.

We proposed an algorithm for Microcalcification detection by introducing shape information. FCS algorithm can detect microcalcification if it's nodular in structure. We have tested112 images from our database. In these images, 95% of microcalcifications were detected correctly and 5% of microcalcifications were failed to detect because of they are not nodular in structure. Thus the research is still being performed to find out the better way to detect microcalcification without nodular structure.

REFERENCES

[1] R.G.Bird,T.W.Wallace,B.C.Yankaskas,"Analysis of cancers missed at screening mammography," Radiology, vol. 184, pp. 613–617,1992

[2] H. Burhenne, L. Burhenne, F. Goldberg, T. Hislop, A.J.Worth, P.M.Rebbeck, and L.Kan, "Interval breast cancers in the screening mammography program of British Columbia: Analysis and classification," Am. J Roentgenol., vol. 162, pp.1067–1071,1994

[3] R.N. Strickland, H.L. Hahn, "Wavelet transforms for detecting microcalcifications in mammograms" IEEE Trans. Med. Imag., vol. 15, no. 2, pp. 218– 229, April 1996

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

[4] H.Yoshida,K Doi, and R. M. Nishikawa, “Automated detection of clustered microcalcifications,” Proc. SPIE (Digital Mammograms Using wavelet Transform Tech., Med.Imag. 1994: Image Process.), vol. 2167, pp. 868–886, February 1994

[5] N. Pandey, Z. Salcic, and J. Sivaswamy, “Fuzzy logic based microcalcification detection”, Proceedings of the IEEE for Signal Processing Workshop, pp. 662– 671,2000

[6] Anne Strauss, Abdeljalil Sebbar, Serge DCsarnaud et al.. "Automatic detection and segmentation of microcalcifications on digitized mammograms", In Proceedings of the Annual International Conference of the IEEE,14(5),pp.1938-1939, 1992

[7] F. Valverde, N. Guil, J. Munoz, Q. Li, M. Aoyama and K. Doi, “A Deformable Model for Image Segmentation in Noisy Medical Images”,International Conference on Image Processing, vol.3, pp. 82-85, Oct. 2001.

[8] K. Doi, H. MacMahon, S. Katsuragawa, R. M.Nishikawa, and Y. Jiang, “Computer-aided diagnosis in radiology: Potential and pitfall,” Eur. JRadiol.,vol. 31, pp. 97–109,1999

[9] Nakayama R,Uchiyama Y,Yamamoto K,Watanabe R, Namba, K, “Computer-aided diagnosis scheme using a filter bank for detection of microcalcification clusters in mammograms”, IEEE Transactions on Biomedical Engineering, Vol 53.No.2,p.273-283, February 2006

[10] R. N. Dave , “Fuzzy shell clustering and application to circle detection in digital images.International Journal on General Systems, 16:343-355,1990

[11] M.N. Gurcan, Y. Yardimci, A.E. Cetin and R.Ansari , “Automated Detection andEnhancement of Microcalcification on DigitalMammograms using Wavelet Transform Techniques”, Dept. of Radiology, Univ.of Chicago,1997

[12] M.Barni, A.Mecocci and G.Perugini, “Application of possibilistic shell-clustering to the detection of craters in real-world imagery”,Proceedings of the IEEE for Geoscience and Remote Sensing Symposium, vol.1, pp.168-170,2000

[13] C.E.Metz,“Some practical issues of experimental design and data analysis in radiological ROC studies,” Invest. Radiol., vol. 24, no. 3, pp. 234–245, Mar. 1989

[14] Hiroyuki Yoshida,“Diagnostic Scheme Using wavelet Transform for Detecting Clustered Microcalcifications in Digital Mammograms”, Acad Radiol, Vol. 3,1996

[15] S.A. Hojjatoleslami and J. Kittler, "Automatic detection of calcification in mammograms", The IEEE Fifth International Conference on Image Processing and Its Applications, pp.139-143. July 1995

[16] Heng-Da Cheng,* Senior Member, IEEE, Yui Man Lui, and Rita I. Freimanis “A Novel Approach to Microcalcification Detection Using Fuzzy Logic Technique”;IEEE transactions on medical imaging, vol. 17, no. 3, June 1998

[17] F. Laine, J. Fan, andW. Yang, “Wavelets for contrast enhancement of digital mammography,” IEEE Eng. Med. Biol. Mag., vol. 14, no. 5, pp.536–550, Oct. 1995.

[18] S. Yu and L. Guan, “A CAD system for the automatic detection of clustered microcalcifications in digitized mammogram film,” IEEE Trans. Med. Imaging, vol.19, no.2, pp.115–126, 1998.

T.Balakumaran received the Bachelors degree in Electronics and Communication Engineering from Bharathiyar University, Coimbatore in 2003 and the Master degree in Applied Electronics from Anna University, Chennai in 2005. Since then, he is working as a Lecturer in Velalar College of Engineering and Technology (Tamilnadu), India. Presently he is a Part time (external) Research Scholar in the Department of Electrical Engineering at Anna University, Coimbatore (India). His fields of interests include Image Processing, Medical Electronics and Neural Networks.

Dr.ILA.Vennila received the B.E Degree in Electronics and Communication Engineering from Madras University, Chennai in 1985 and ME Degree in Communication System from Anna University, Chennai in 1989. She obtained Ph.D. Degree in Digital Signal Processing from PSG Tech, Coimbatore in 2006. Currently she is working as Assistant Professor in EEE Department, PSG Tech and her experience started from 1989; she published about 35 Research Articles in National, International Conferences National and International journals. Her area of interests includes Digital Signal Processing, Medical Image processing, Genetic Algorithm and fuzzy logic.

Mr.C.Gowri Shankar received the B.E. Electrical and Electronics Engineering from Periyar University in 2003 and M.E Applied Electronics from Anna University, Chennai in 2005. Since 2006, he has been a Ph.D. candidate in the same university. His research interests are Multirate Signal Processing, Computer Vision, Medical Image Processing, and Pattern Recognition. Currently, he is working in Dept of Electrical and Electronics Engineering, Velalar College of Engineering and Technology, Erode.

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

The Fast Haar Wavelet Transform for Signal & Image Processing

V.Ashok
Department of BME,
Velalar College of Engg.&Tech.
Erode, India – 638012.

T.Balakumaran
Department of ECE
Velalar College of Engg.&Tech
Erode, India – 638012

C.Gowrishankar
Department of EEE
Velalar College of Engg.&Tech
Erode, India – 638012

Dr.ILA.Vennila
Department of ECE,
PSG College of Technology,
Coimbatore,
TamilNadu, India

Dr.A.Nirmal kumar
Department of EEE,
Bannari Amman Institute of Technology,
Sathyamangalam,
TamilNadu, India

***Abstract*- A method for the design of Fast Haar wavelet for signal processing & image processing has been proposed. In the proposed work, the analysis bank and synthesis bank of Haar wavelet is modified by using polyphase structure. Finally, the Fast Haar wavelet was designed and it satisfies alias free and perfect reconstruction condition. Computational time and computational complexity is reduced in Fast Haar wavelet transform.**

Keywords- computational complexity, Haar wavelet, perfect reconstruction, polyphase components, Quardrature mirror filter.

I. INTRODUCTION

The wavelet transform has emerged as a cutting edge technology, within the field of signal & image analysis. Wavelets are a mathematical tool for hierarchically decomposing functions. Though routed in approximation theory, signal processing, and physics, wavelets have also recently been applied to many problems in computer graphics including image editing and compression, automatic level-of-detailed controlled for editing and rendering curves and surfaces, surface reconstruction from contours and fast methods for solving simulation problems in 3D modeling, global illumination and animation [1].

Wavelet theory was developed as a consequence in the field of study the multi-resolution analysis. Wavelet theory can determine the nature and relationship of the frequency and time by analysis at various scales with good resolutions.

Time-Frequency approaches were obtained with the help of Short Time Fourier Transform (STFT). For the better time (or) frequency resolution (but not both) can be determined by individual preference (or) convenience rather than by necessity of the intrinsic nature of the signal, the wavelet analysis gives the better resolution [2].

According to the applications, the biomedical researchers have large number of wavelet functions from which to select the one that most closely fits to the specific application. Wavelet theory has been successfully applied to a number of biomedical problems [3-5]. Many applications such as image compression, signal & image analysis are dependent on power availability. In this paper, a method for design of Haar wavelet for low power application is proposed. The main idea of this proposed method is the decimated wavelet coefficients are not computed. This makes the conservation of power and reduces the computation complexity. The Haar wavelet which makes the low power design is simple and fast. The proposed design approach introduces more savings of power.

This paper organised as follows. In Section II, the existing Haar wavelet is introduced. In section III presents Haar wavelet analysis bank reduction. In section IV presents Haar wavelet synthesis bank reduction. In section V presents Haar wavelet and Fast Haar wavelet experimental results are shown as graphical output representation to the signal and image processing and we conclude this paper with section VI.

II. HAAR WAVELET STRUCTURE

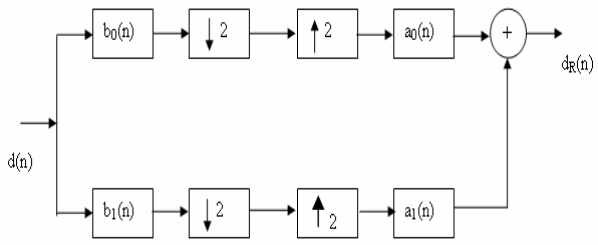


Fig. 1 Two channel wavelet structure

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

The wavelet transform can be implemented by a two channel perfect reconstruction (PR) filter bank [6]. A filter bank is a set of filters, which are connected by sampling operators. Fig.1 shows an example of a two-channel filter bank applied by one dimensional signal. d*(n)* is an input signal and $d_R(n)$ is reconstructed signal. In the analysis bank, $b_0(n)$ is a analysis low pass filter and $b_1(n)$ is a analysis high pass filter. However in practice, the responses overlap, and decimation of the sub-band signals, which are results in aliasing. The fundamental theory of the QMF bank states that the aliasing in the output signal $d_R(n)$ can be completely canceled by the proper choice of the synthesis bank [7]. In the synthesis bank, $a_0(n)$ is the reconstruction low pass filter(LPF) and $a_1(n)$ is the reconstruction high pass filter (HPF). Low pass analysis coefficients of Haar Wavelet is $\left[\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}}\right]$. High pass analysis coefficients of Haar Wavelet is $\left[-\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}}\right]$. Low pass synthesis coefficients of Haar Wavelet is $\left[\frac{1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}}\right]$. High pass synthesis coefficients of Haar Wavelet is $\left[\frac{1}{\sqrt{2}} \quad -\frac{1}{\sqrt{2}}\right]$.

III. HAAR WAVELET ANALYSIS BANK REDUCTION

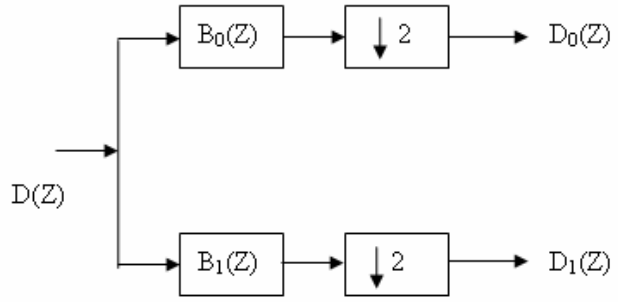


Fig. 2 Analysis bank of wavelet structure

Fig.2 shows analysis bank of wavelet structure. d(n) is an input signal, $d_0(n)$ is an low pass output of d(n) and $d_1(n)$ is high pass output of input signal.

For simplicity write in Z domain

$$D_0(Z) = \tfrac{1}{2}[D(Z^{1/2})\,B_0(Z^{1/2}) + D(-Z^{1/2})B_0(-Z^{1/2})] \quad (1)$$

$$D_1(Z) = \tfrac{1}{2}[D(Z^{1/2})\,B_1(Z^{1/2}) + D(-Z^{1/2})B_1(-Z^{1/2})] \quad (2)$$

At Perfect Reconstruction condition, No Aliasing Components presents

$$D_0(Z) = \tfrac{1}{2}[D(Z^{1/2})B_0(Z^{1/2})] \quad (3)$$

$$D_1(Z) = \tfrac{1}{2}[D(Z^{1/2})B_1(Z^{1/2})] \quad (4)$$

From Quadrature Mirror Filter by [7], analysis filters are chosen as follows

$$B_0(Z) = B(Z) \leftrightarrow b(n) \quad (5)$$

$$B_1(Z) = B(-Z) \leftrightarrow (-1)^n b(n) \quad (6)$$

Transfer function B(Z) of an LTI system can decomposed into its polyphase components[9] .

B(Z) can be decomposed into

$$B_0(Z) = \sum_{\lambda=0}^{M-1} Z^{-\lambda} B_\lambda(Z^M) \quad (7)$$

In Haar Wavelet M=2

So Low pass filter & High pass filter is

$$B_0(Z) = B_{00}(Z^2) + z^{-1}B_{01}(Z^2) \quad (8)$$

$$B_1(Z) = B_{00}(Z^2) - z^{-1}B_{01}(Z^2) \quad (9)$$

Sub $B_0(Z)$, $B_1(Z)$ in Eq (3) & (4)

$$D_0(Z) = \tfrac{1}{2}[D(Z^{1/2})(B_{00}(Z) + z^{-1/2}B_{01}(Z))]$$

$$D_0(Z) = \tfrac{1}{2}[D(Z^{1/2})(B_{00}(Z) + \tfrac{1}{2}z^{-1/2}D(Z^{1/2})B_{01}(Z)] \quad (10)$$

In Haar wavelet $B_{00}(Z) = B_{01}(Z)$

$$D_0(Z) = B_{00}(Z)[\tfrac{1}{2}D(Z^{1/2}) + \tfrac{1}{2}Z^{-1/2}D(Z^{1/2})] \quad (11)$$

Like

$$D_1(Z) = D(Z^{1/2})B_{00}(Z) - \tfrac{1}{2}Z^{-1/2}D(Z^{1/2})B_{01}(Z)$$

$$D_1(Z) = B_{00}(Z)[\tfrac{1}{2}D(Z^{1/2}) - \tfrac{1}{2}Z^{-1/2}D(Z^{1/2})] \quad (12)$$

Combining Eq (11) & (12)

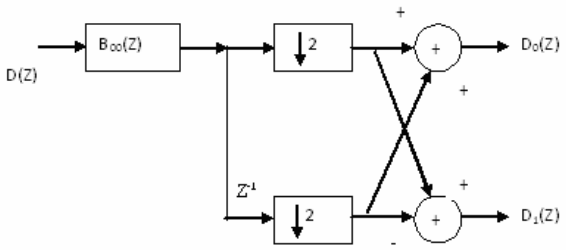


Fig. 3 Modified analysis bank structure

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

$B_{00}(Z) \longleftrightarrow b_{00}(n)$. In haar wavelet $b_{00}(n) = \frac{1}{\sqrt{2}}$.

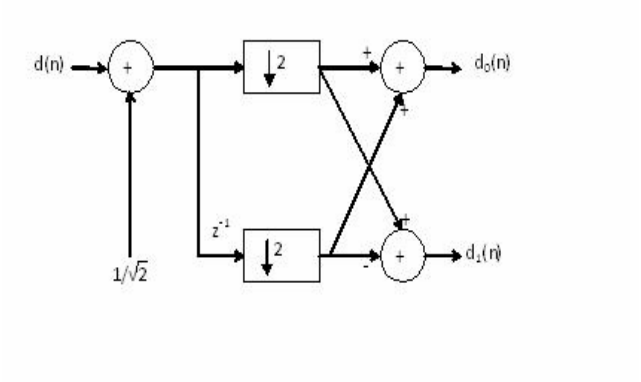


Fig. 4 Fast Haar wavelet analysis bank

Shifting the down sampler to the input bring reduction in the computational complexity of factor 2 along with it. Fig.4 shows Fast Haar wavelet analysis structure compared to original Haar wavelet structure, Number of arithmetic calculations are reduced in Fast Haar wavelet structure. But using above method computational complexity [10] reduced in less than quarter of original computational complexity.

IV. HAAR WAVELET SYNTHESIS BANK REDUCTION

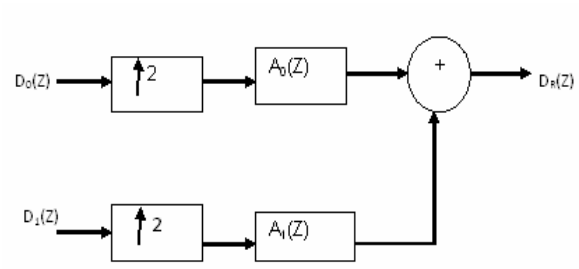


Fig. 5 Synthesis bank of wavelet structure

Fig.5 shows synthesis bank of wavelet structure. $d_0(n)$ is low pass input signal, $d_1(n)$ is high pass input signal and $d_R(n)$ is reconstructed signal

For simplicity write in Z domain

$$D_R(Z) = A_0(Z)D_0(Z^2) + A_1(Z)D_1(Z^2) \quad (13)$$

From Quadrature Mirror Filter by [8] at perfect reconstruction, filters are chosen as follows

$$A_0(Z) = 2B(Z) \leftrightarrow 2b(n) \quad (14)$$

$$A_1(Z) = -A(-Z) = -2B(-Z) \leftrightarrow 2(-1)^{n+1}b(n) \quad (15)$$

Refer to Eq (7)

A(Z) is decomposed into

$$A(Z) = \sum_{\lambda=0}^{M-1} Z^{-\lambda} A\lambda(Z^M) \quad (16)$$

In Haar Wavelet M=2

$$A_0(Z) = A_{00}(Z^2) + z^{-1}A_{01}(Z^2) \quad (17)$$

$$A_1(Z) = -A_{00}(Z^2) + z^{-1}A_{01}(Z^2) \quad (18)$$

Sub Eq. 17 & 18 in (13)

$$D_R(Z) = D_0(Z^2)[A_{00}(Z^2) + z^{-1}A_{01}(Z^2)] + [-A_{00}(Z^2) + z^{-1}A_{01}(Z^2)]D_1(Z^2)$$

$$D_R(Z) = A_{00}(Z)[D_0(Z^2) - D_1(Z^2)] + z^{-1}A_{01}(Z^2)[D_0(Z^2) + D_1(Z^2)] \quad (19)$$

Up sampler at the input of the synthesis filter bank will moved to output. So Eq.(19) can be drawn by

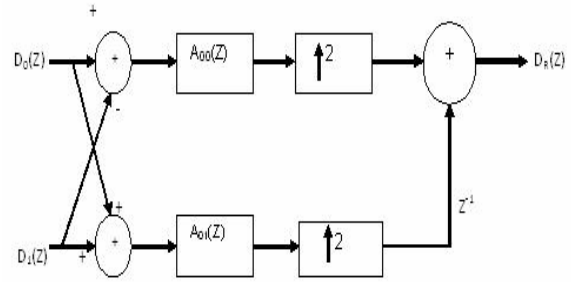


Fig. 6 Modified synthesis bank structure

In Haar wavelet $A_{00}(Z) = A_{01}(Z) = B_{00}(Z)$

In Haar wavelet $b_{00}(n) = a_{00}(n) = \frac{1}{\sqrt{2}}$.

Draw in time domain

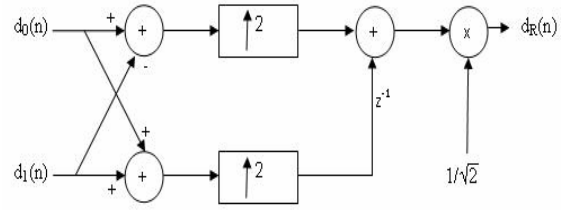


Fig. 7 Fast Haar wavelet synthesis bank

Combining Fig.4 & Fig.7, Fast Haar Wavelet Structure is obtained. Compared to Fig.2, Number of Mathematical calculations are reduced in Fast Haar Wavelet Structure is shown in Fig.8.

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

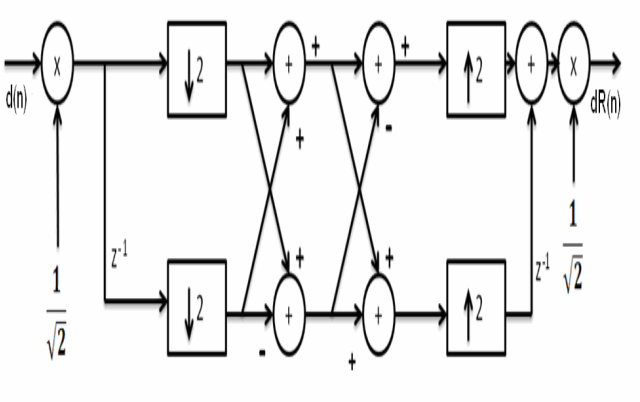


Fig. 8 Fast Haar wavelet structure

V. EXPERIMENTAL RESULTS

The results of applying, for one subject, which the signal is taken from laser based noninvasive Doppler indigenous developed equipment, the novel Fast Haar wavelet with approximation data are shown in Fig.9 shows that difference between original haar wavelet and Fast haar wavelet are matched well. The Error rate between existing and proposed Fast Haar wavelet at -90dB are shown in Fig. 11.

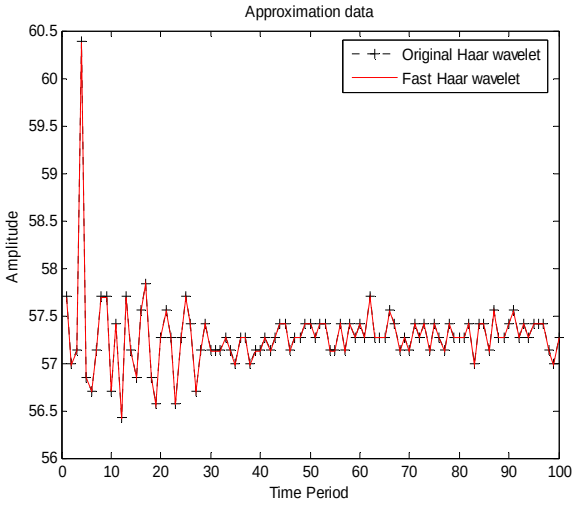


Fig. 9 Results of approximation data compared to existing and Proposed Fast Haar wavelet Transform.

Similarly from the same novel Fast Haar wavelet with detail data are shown in Fig.10 shows that difference between original Haar wavelet and Fast Haar wavelet are matched well. The Error rate between existing and proposed Fast Haar wavelet at -160dB to -220dB are shown in Fig.11.

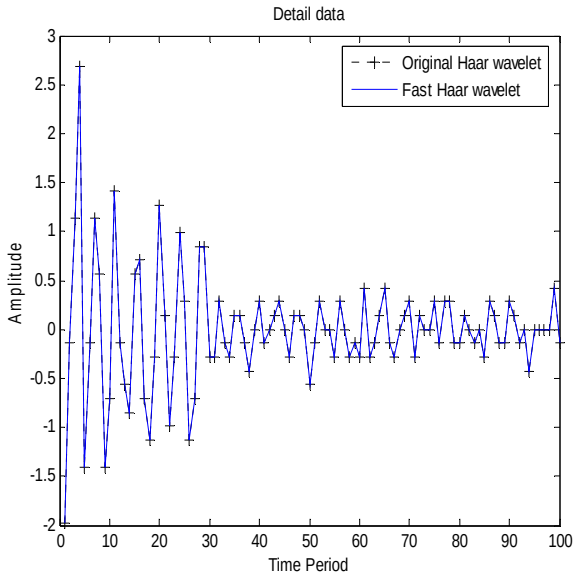


Fig. 10 Results of detail data compared to existing and Proposed Fast Haar wavelet Transform.

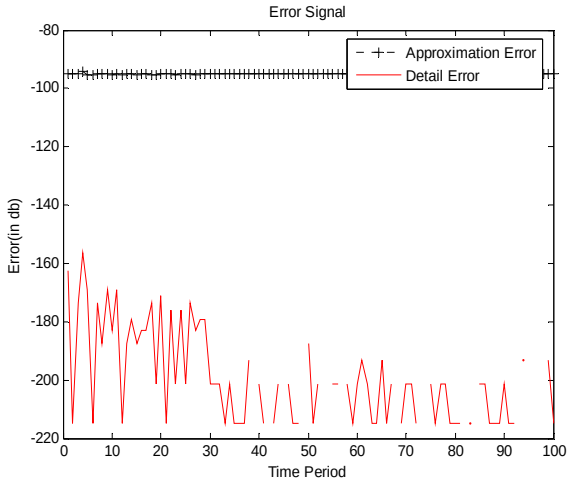


Fig. 11 Results of Error rate compared to existing and Proposed Fast Haar wavelet Transform

We have checked our proposed method in image processing also. Lowpass output was obtained by applying original Haar wavelet and proposed Fast Haar wavelet. Fig.12(a) shows Lena image, Fig.12(b) shows lowpass image of lena by applying original Haar wavelet transform and Fig.12(c) shows lowpass image by applying Fast Haar wavelet transform. Fig.12(d) shows difference between Fig.12(b) & Fig.12(c) From the Fig.12(d), it is clearly visible difference value for all coefficients are less.

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

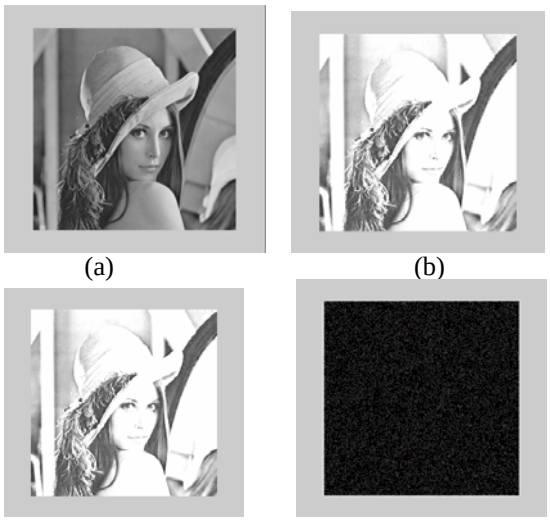


Fig.12 Comparison of Fast haar wavelet with original Haar wavelet
a) Lena image (b) Lowpass of Lena image by original Haar wavelet
(c) Lowpass of Lena image by Fast Haar wavelet
(d) Difference between lowpass output by original Haar wavelet & Fast Haar wavelet

VI. CONCLUSION

This work presents a novel Fast Haar wavelet estimator, for application to biosignals such as noninvasive doppler signals and medical images. . In this paper, signals and images are decomposed and reconstructed by Haar wavelet transform without convoution. The proposed method allows for the dynamic reduction of power and computational complexity than the conventional method.The error rate between the conventional and the proposed method was reduced in the signal and image procesing.

REFERENCES

[1] Eric J.stollnitz, Tony D.Derose and david H.Salesin, Wavelets for computer graphics – theory and applications book, Morgan kaufmann publishers, Inc.San Francisco California

[2] O. Rioul and M. Vetterli , 'Wavelets and Signal Processing', IEEE Signal Processing Mag, pp 14–18, Oct 1991.

[3] Fig.liola and E. Serrano, 'Analysis of physiological time series using wavelet transforms,' IEEE Eng. Med. Biol, pp 74 – 80, May/June 1997.

[4] Aldroubi and M.A. Unser, Eds , Wavelets in Medicine and Biology.

[5] P. Salembier, 'Morphological multiscale segmentation for image coding,' Signal Process, vol. 38, pp. 359–386, 1994.

[6] P.P. Vaidyanathan , 'Multirate Systems and Filter Banks,' Englewood Cliffs, NJ: Prentice-Hall, 1993.

[7] P.P.Vaidyanathan; Quadrature Mirror filter banks, M band extensions and Perfect Reconstruction Techniques. IEEE ASSP Magazine, Vol 4, pp 4-20, July 1987.

[8] J.H.Rothweiler: polyphase, Quadrature filters – A new subband coding technique. IEEE ICASSP'83, pp.1280-1283, 1983

[9] R.E.crochiere,L.R.Rabiner: Multirate Digital signal processing. Englewood Cliffs:prientice hall,1983.

[10] M.J.T Smith, T.P.Barnwell III: A Procedure for designing exact Reconstruction filter banks for tree structured Sub-band Coders. Proc IEEE ICASSP'84, pp.27.1.1-27.1.4, March 1984.

AUTHORS PROFILE

Mr.V.Ashok received the Bachelors degree in Electronics And Communication Engineering from Bharathiyar University, Coimbatore in 2002 and the Master degree in Process Control And Instrumentation Engineering form Annamalai University, Chidambaram in 2005. Since then, he is working as a Lecturer in Velalar College of Engineering and Technology (Tamilnadu), India. Presently he is a Part time (external) Research Scholar in the Department of Electrical Engineering at Anna University, Chennai (India). His fields of interests include Medical Electronics, Process control and Instrumentation and Neural Networks.

Mr.T.Balakumaran received the Bachelors degree in Electronics and Communication Engineering from Bharathiyar University, Coimbatore in 2003 and the Master degree in Applied Electronics from Anna University, Chennai in 2005. Since then, he is working as a Lecturer in Velalar College of Engineering and Technology (Tamilnadu), India. Presently he is a Part time (external) Research Scholar in the Department of Electrical Engineering at Anna University, Coimbatore (India). His fields of interests include Image Processing, Medical Electronics and Neural Networks.

Mr.C.Gowri Shankar received the B.E Electrical and Electronics Engineering from Periyar University in 2003 and M.E Applied electronics from Anna University, Chennai in 2005. Since 2006, he has been a Ph.D. candidate in the same university. His research interests are Multirate Signal Processing, Computer Vision, Medical Image Processing, and Pattern Recognition. Currently, he is working in Dept of Electrical and Electronics Engineering, Velalar College of Engineering and Technology, Erode.

Dr.ILA.Vennila received the B.E Degree in Electronics and Communication Engineering from Madras University, Chennai in 1985 and ME Degree in Communication System from Anna university, Chennai in 1989. She obtained Ph. D. Degree in Digital Signal Processing from PSG Tech, Coimbatore in 2006. Currently she is working as Assistant Professor in EEE Department, PSG Tech and her experience started from 1989; she published about 35 Research Articles in National, International Conferences National and International journals. Her area of interests includes Digital Signal Processing, Medical Image processing, Genetic Algorithm and fuzzy logic

Dr.A.Nirmalkumar. A, received the B.Sc.(Engg.) degree from NSS College of Engineering, Palakkad in 1972, M.Sc.(Engg.) degree from Kerala University in 1975 and completed his Ph.D. degree from PSG Tech in 1992. Currently, he is working as a Professor and Head of the Department of Electrical and Electronics Engineering in Bannari Amman Insititute of Technology, Sathyamangalam, Tamilnadu, India. His fields of Interest are Power quality, Power drives and control and System optimization.

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

A Survivability Strategy in route optimization Mobile Network by memetic algorithm

K .K. Gautam
Department of Computer Science & Engineering
Roorkee Engineering & Management Technology Institute,
Shamli (247774) India

Anurag Rai
Department of Information Technology
College of Engineering Roorkee
Roorkee (247667) India

***Abstract*— The capability to provide network service even under a significant network system element disruption is the backbone for the survival of route optimize of mobile network Technology in today's world. Keeping this view in mind, the present paper highlights a new method based on memetic algorithm .**

Keywords- Survivability, Mobile Network, memetic algorithm., PAN,.

I. INTRODUCTION

Network survivability is considered to cope, it with increasing demand for reliable network system. Network survivability for the route optimization is an essential aspect of reliable communication service. Survivability consists not only of robustness against failure occurring due to natural faults. In mobile networks infrastructure element such as base station (BS), base station Controller (BSC), wired links, and mobile switch centre (MSC), are employed to provide and maintain essential services, hence the operation interruption of a network component affects overall or partial network services. Wireless access network have unique characteristics to support mobile users, which can result in different survivability and security aspect [3]. Therefore wireless survivability strategies must be designed to improve the service available rate of the network component system (4), (3).

Mobile user authentication is necessary when a mobile user wants to request service provided by the service providers survivable (SPS) in the visited domains. In this paper, we present a survivability strategy in mobile networks method by the use of memetic algorithm.

A network could be as simple as a forum held in a city between people, where people use the opportunity to communicate with each other, they use a network by the use of memetic algorithm has the potential for setup the survivability. Fundamental to distribute mechanics is the effect of measurement on a state. If some property of a general state is measured, it collapses to an eigenstate of the property and cannot be 'rebuilt' into the original state. Information can be encoded into a general optimize set up. Mobile networks can have very complex form of hierarchy e.g. Mobile networks in a mobile network visiting mobile nodes (VMNS) in mobile networks and so on. This situation is called as a nested mobile network.

Many important problems arising in science, industry and commerce, mobile networks fall very neatly into the read-made category of optimization problem.

II. SURVIVABILITY

Traditional security research is primarily focused on the detection and prevention of intrusion and attacks rather than on continued correct operation while under attack. Fault tolerance is usually concerned with redundancy that is required to detect and correct up to a given number of naturally occurring faults. Nature is not malicious and conventional failure model make significant assumptions, in particular, assuming faults to be independent and random. The presence of intelligent adversarial attacks can protocol vulnerability often become more important considerations in the presence of an adversary[1,2]

There are a number of definitions of survivability. The one we use here is from the Software Engineering Institute, which emphasizes timeliness, survivability under attack and failure, and that detection of attack is a vital capability.

Survivability is the capability of a system to fulfill its mission in a timely manner, even in the presence of attacks or failures. Survivability goes into the demon of security and fault. Tolerance of focus on delivery of essential service even when system is entered or experiences failures, and rapid recovery of full service, when conditions improve. Unlike traditional security measures that require central control and administrative, survivability addresses highly distributed unbounded network environment that lacks central control and unified security policies. Mobile networks can have very complex form of hierarchy e.g. Mobile networks in a mobile network visiting mobile nodes (VMNS) in mobile networks and so on.

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

III. The Three Rs: Resistance, Rrcognition, And Recovery

The focus of survivability is on delivery of essential services and preservation of essential assets. Essential service and asserts are those system capabilities that are critical o fulfilling mission objectives. Survivability depends on three key capabilities: resistance, recognition, and recovery. Resistance is the capability to detect attacks as they occur and to evaluate the extent of damage and compromise. Recovery, a hallmark of survivability is the capability to maintain essential service and asserts during attacks, limit the extent of damage and restore full service following attack.

We further extend this definition to require that survivability system be able to quickly incorporate lesion learned from failure, evolve, and adapt to emerging threats. We call this survivability feature refinement.

We can classify survivable mobile wireless networking requirement into four categories based on (2): (i) resistance requirement; (ii) recognition requirement; (iii) recovery requirements; and (iv) refinement requirement. We can also describe a requirement definition process (3). This includes the definition of system and survivability requirement, legitimate and intruder usage requirement, development requirement, operation requirement, and evolution requirement. Essential service must be identified and specified for the penetration, exploration, and exploitation phases of the attack.

The approach has guided this work and is recommended for more detailed requirement analyses for future mobile wireless network.

Ultimately, there are two distinct aspects of survivability that apply at all networking layers.

IV. Information Access Requirement

Does the user have access to the information or service required to complete the task in the presence of failure or attack? For eg. it is possible to replicate service or information and provide them locally when the network gets partitioned. However End-to-end communication should not be mandated in these cases.

V. End-To-End Communication

On the other hand there are interactive application , inter-personal communication such as voice calls, or dynamically generated information such as current sensor data, which require true end–to–end connectivity. Do existing sessions survive? Can the user create new session to reach the intended communication end- point even in the presence of failures and attacks? This requires that the communication end–point themselves survive and that the communication end–points themselves survive and that the adversary must not be able to permanently partition the network. Furthermore, the adversary must not be able to permanently disable access to required services such as authentication, naming, resource discovery, or routing.

VI. Mobile Network Survivability

Existing work on survivability in the context of cellular telephone networks concentrates primarily on infrastructure survivability (for e.g. see the outage index metrics (80) and does not consider adversarial attacks. However, they offer some insight on quantifying survivability and the role of network management tools.

Networks are vulnerable during upgrades, especially those involving software. Furthermore, rapid evolution leads to learning – cure problems as well as – over – concentration leads or service into single points of failure. This problem is exacerbated by deficits in network management tools to operate and maintain increasingly complex system. Architectural improvement applicable to mobile include the use of redundant networks

.

VII. Base Station

In more environment, a cell that is geographical region unit is covered by the radio frequency of a base station. Each call is controlled by a BS which has a fixed connection to a BSC (or RNC). In mobile network infrastructure element such as base station controller (BSC), wired links and mobile switch centre (MSC) are employed to provide and maintain essential service. Hence the operation interruption of a network component affects overall or partial network services.

A radiation antenna is classified as omni directional and directional with an ommni directional antenna, a single frequency spreads out in all directions of 360 coverage. A cell is directional antenna with each different set channel.

VIII. System State Of Base Station

The BS system, including antenna parts cannot provide partial or whole service function for coverage cell when single or more fatal failures occur in the BS system. In this paper, we consider that system failures are caused by key distribution method. For example, by interrupt sequence mishandling, overall system operation falls into failure state because of unanticipated handled interruption to a component of the system.

IX. Personal Area Network

A mobile network can have a hierarchical structure e.g. mobile network within another mobile network. This situation is referred to as nested mobile network. A personal area network (PAN) may travel a vehicle, which also contains a mobile network of larger on scale, fig 1 illustrate a simple

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

larger scale. MR-1, MR-2 are attached to their own home links. A wireless personal area network moves as a single unit with one or more mobile routers that connect it to global internet.

X. MULTI OBJECTIVE OPTIMIZATION (MOO)

An unaccompanied multi objective optimization problem is an example of route optimization for mobile network. Because mobile moves as a single unit with one or more mobile routers that connect it to the global internet[9,11] We defined this problem as

"Minimize" z = f(x)

Where f(x) = (f1(x), f2(x)…..fn(x))

Subject to x∈x

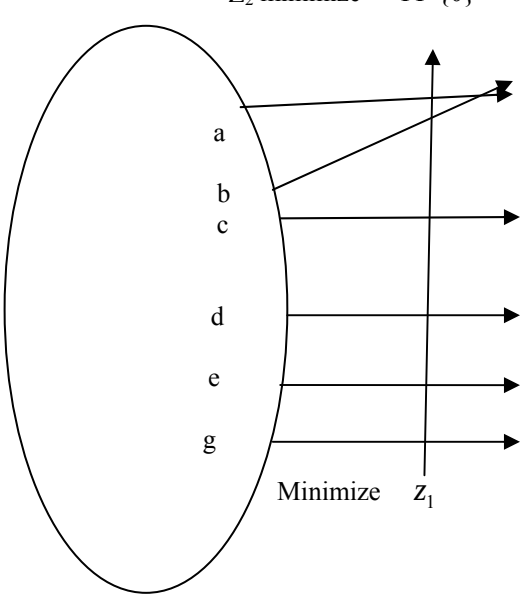


fig-1

Fig-1 an example of multi objective optimization problem with mobile search space (MSS) x, as vector fitness function f that maps solution in x to objective vector made up of two component (mobile routers) 'costs' z_1 and the z_2 minimized.

Here if we define

a=mobile router -1

b=mobile router-2

.

.

.

.

And

1=access router -1

2=access router-2

.

.

.

.

This fig is also defined a routing inefficiency for the traffic management and designed an imported rout optimization schemes for traffic management of mobile networks. Thes concept of traffic management for the network mobility was introduce the signaling over heads of a number of hosts moving as a group as MRs.

XI. MEMETIC ALGORITHM APPROACH

The impressive record of Memetic Algorithms producing high quality solution in combinalarial optimization and in real – world application (e.g. see page 220[5]) is sometimes cited as a testament of their inherent effectiveness of robustness as black box search. However, since the advent of the no free lunch theorems (6,7,8) we know that MA;S like any other search algorithm, are only reality good to the extent to which they can be "aligned" to the specific features of a route optimization problems in mobile networks. None the less, MAs, like therefore bears evolutionary algorithms (EAs), do have unassailable advantage over other more traditional search techniques: that is their flexibility. This flexibility has important advantage, as has to solve mobile route optimization problems: one is to choose some traditional techniques. And them simplify or otherwise other the problems.

As in any other optimization scenario as route optimization problems, we should know that the out set what is a desirable outcomes; the Memetic Algorithm frame work proposed above required. The operators and procedures be selected based on their current success abs. When a mobile network moves from one place to another it change its point attachment to the internet, which also makes changes to its reach ability and to the internet topology.[9,10,11]

XII. PERFORMANCE MEASURES IN MAS FOR MOO

If one is developing or using an algorithm for optimization it almost goes without saying that there should be some way to measures its performance. In MOO the situation is the same regarding the time aspect of performance assessment but the quality aspect is clearly more difficult. the extensive array of existing met heuristic, issues and methods reviewed in the section above gives a richer basis from which to design new MAs than do the existing MAs for MOO themselves. In a typical cellular network, the area of coverage is after geographically divided into hexagonal cells. The call is the basic unit of a cellular system.

In recent years, Muscat and Krasnogor have provided a guiding manifesto for putting the "Memetic" back in Memetic algorithm (9,10) advocating.

Candidate MA framework for MOO

- MN: = initialize(MN)
- A: = Nondom (MN)

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

Algorithm Candidate MA framework for MOO

- MN: = Initialize (MN)
- MN: = Nondom (MN)
- **while** stop_criterion not satisfied **do**
- **while** stagnation _criterion not satisfied **do**
- SAMN:=SelectFrom(PUA,sel_sc hed(succ(SEL)))
- SAMN:=Vary(SAMN,var_sched (succ(VAR)))
- SAMN":=LocalSearch(SAMN",I s_sched(succ(LS)))
- MN:=Replace(PUC",rep_sched(s ucc(RED)))
- A:=Reduce(Nandom(AUSAMN),red_sched (succ(RED)))
- **end while**
- MN:=Randomlmmigrants(P,imm_sched(succ(IMM)))
- **return** (A)

Here we represent an Algorithm, we put forward a simple framework that could serve as a guide for making a more Memetic MA for MOO. In line1, MN (Mobile Networks) of solution is initialized. As usual this procedure may be simply random or it may employ some heurists (s). Line 2 sets the archive A to the no dominated solution from MN. Thereafter, the main loop of the MA begins line 4 sets up an inner loop in which a stagnation criterion is checked. This should be based on some memeplex which monitors progress in diversity, proximity, and /or some other criteria. Line5-9 gives a very high level description of the update of the MN and archive. Five different 'schedulers' are employed, basically corresponding to mating selection, reproduction, lifetime learning, survival selection and update of the archive, respectively. Each scheduler chooses from a memeplex of operators, based on estimates of the current success of those operators. E.g. in line 5, Select from is the operation of mating selection, the domain of which is the union of the MN and archive, and co-domain is a Small Area Mobile Networks (SAMN), the selection is controlled by the scheduler, sel_sched, which use a success measures to choose one operators for the set SEL, of currently available operators for selection. Notice that MN and A are potentially of variable size, in this scheme. In line 11, the MN is updated using some immigration policy to rerelease it from stagnation, the archives of no dominated solution are returned in line 13.

The frame work proposed is rather broad and actually instantiating it requires us to consider how we should resolve many choices, including those considered in the following sections, at the very least?

CONCLUSION

In this paper, we have proposed a scheme for mobile service use of BS system and memetic alogithm. The survavility of Route optimization scheme in nested mobile network modifying the process of Memetic Algorithm . And hence the basic support protocol for survivability of Route optimization scheme for mobile network needs to be extended with an appropriative route optimization scheme. we proposed scheme can achieve the mobile route optimization environment, it may get a survivability scheme.

ACKNOWLEDGMENT

The author would like to thank Dr. H.N. Dutta Director, REMTech for his moral support in carrying out the work

REFERENCES

[1] J.Kabara , P. Krishna Murthy, and D. Tipper, "Information auuurance in wireless network". In proc. IEEE workshop on DIREN'02.

[2] U. Varshney A.P Snow and A.D. Malloy, "Designing Survivable wireless and mobile network" in proc. IEEE WCNC'99, neworeleans, LA, Sep. 1994, pp.30-34.

[3] D.Tipper, S. Rammaswamy, and T. Dahiberg,"pcs network survivability" in proc. IEEE WCNC, new or leans LA, sep 1999, invited paper.

[4] D. Samfat, R. molva, N. Asokan, "Untraceability in mobile Network" Proceeding of Mobi COM'95, Berkely, November 1995.

[5] Zhang Bin, Wujing-Xing Proc. Of the feb.2003 ICCNMC'03 IEEE.

[6] Sangjoon Park, Jiyoung Song, Byunaggi Kim, IEEE Trans of Veh. Tech. Vol. 55 pp 328-339.

[7] Ashotosh Dutta , James Burns , K. Daniel Wong, Ravi jain, Ken Young Telcordia Technologies, 445 South Street, Morristown, NJ 07960 pp 1-6

[8] Sangjoon Park , Jiyoung song, and Byunggi Kim, Member, IEEE "A Survivability Strategy in Mobile Networks. IEEE TRANSACTION ON TECHNOLOGY, VOL. 55,NO.328-340.

[9] M.A. Abido, A new multiobjective evolutionary algorithm for environmental /economic power dispatch. In Power Engineering Society Summer Meeting, Vol. 2, page 1263-1268, IEEE, 2001.

[10] Pragya Lamsal, "Network Mobility", Research Seminar on Hot Topics in Internet Protocols, page 1-2.

[11] Joshua Knowles and David Corne, "Memetic algorithms for multiobjective optimization: issues, methods and prospect, page 1-40.

AUTHORS PROFILE

Authors Profile .. K K Gautam is the Dean in the Roorkee Engineering & Management Technology Institute, Shamli-247 774, India.

Anurag Rai is the head of the Information Technology Department in College of Engineering Roorkee, Roorkee- 247 667, India.

Analysis of Large-Scale Propagation Models for Mobile Communications in Urban Area

M. A. Alim*, M. M. Rahman, M. M. Hossain, A. Al-Nahid
Electronics and Communication Engineering Discipline,
Khulna University,
Khulna 9208, Bangladesh.

*Corresponding author.

***Abstract*— Channel properties influence the development of wireless communication systems. Unlike wired channels that are stationary and predictable, radio channels are extremely random and don't offer easy analysis. A Radio Propagation Model (RPM), also known as the Radio Wave Propagation Model (RWPM), is an empirical mathematical formulation for the characterization of radio wave propagation as a function of frequency. In mobile radio systems, path loss models are necessary for proper planning, interference estimations, frequency assignments and cell parameters which are the basic for network planning process as well as Location Based Services (LBS) techniques. Propagation models that predict the mean signal strength for an arbitrary transmitter-receiver (T-R) separation distance which is useful in estimating the radio coverage area of a transmitter are called large-scale propagation models, since they characterize signal strength over large T-R separation distances. In this paper, the large-scale propagation performance of Okumura, Hata, and Lee models has been compared varying Mobile Station (MS) antenna height, Transmitter-Receiver (T-R) distance and Base Station (BS) antenna height, considering the system to operate at 900 MHz. Through the MATLAB simulation it is turned out that the Okumura model shows the better performance than that of the other large scale propagation models.**

Keywords- Path Loss; Okumura model; Hata model; Lee model;

I. INTRODUCTION

In mobile radio systems, path loss models are necessary for proper planning, interference estimations, frequency assignments, and cell parameters which are basic for network planning process as well as LBS techniques that are not based on GPS system [3]. A Radio Propagation Model (RPM), also known as the Radio Wave Propagation Model (RWPM) or the Radio Frequency Propagation Model (RFPM), is an empirical mathematical formulation for the characterization of radio wave propagation as a function of frequency, distance and other conditions. Most radio propagation models are derived using a combination of analytical and empirical methods. In general, most cellular radio systems operate in urban areas where there is no direct line-of-sight path between the transmitter and receiver and where the presence of high rise buildings causes severe diffraction loss.

Propagation models that predict the mean signal strength for an arbitrary transmitter-receiver seperation distance are useful in estimating the radio coverage area of a transmitter and are called large-scale propagation model. On the other hand, propagation models that characterize the rapid fluctuations of the received signal strength over very short travel distances or short time durations are called small scale or fading models [1].

In this paper, the wideband propagation performance of Okumura, Hata, and Lee models has been compared varying Mobile Station (MS) antenna height, propagation distance, and Base Station (BS) antenna height considering the system to operate at 900 MHz. Through the MATLAB simulation it turned out that the Lee model outperforms the other large scale propagation models.

II. LITERATURE REVIEW

Path loss characteristics of a channel are commonly important in wireless communications and signal propagation. Path loss may occur due to many effects, such as free-space loss, refraction, diffraction, reflection, aperture-medium coupling loss and absorption. Path loss is also influenced by terrain contours, environment (urban or rural, vegetation and foliage), propagation medium (dry or moist air), the distance between the transmitter and the receiver, and the height of antennas [4].

Path loss normally includes propagation losses caused by

- The natural expansion of the radio wave front in free space (which usually takes the shape of an ever-increasing sphere)
- Absorption losses (sometimes called penetration losses)
- When the signal passes through media not transparent to electromagnetic waves, diffraction losses when part of the radiowave front is obstructed by an opaque obstacle and
- Losses caused by other phenomena.

The signal radiated by a transmitter may also travel along many and different paths to a receiver simultaneously; this effect is called multipath. Multipath can either increase or decrease received signal strength, depending on whether the

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

individual multipath wavefronts interfere constructively or destructively.

In wireless communications, path loss can be represented by the path loss exponent, whose value is normally in the range of 2 to 4 (where 2 is for propagation in free space, 4 is for relatively lossy environments. In some environments, such as buildings, stadiums and other indoor environments, the path loss exponent can reach values in the range of 4 to 6. On the other hand, a tunnel may act as a waveguide, resulting in a path loss exponent less than 2 [4].

The free-space path loss is denoted by $L_p(d)$, which is

$$\overline{L}_p(d) = -20\log_{10}\left(\frac{c/f_c}{4\pi d}\right) \text{(dB)}$$

where, c = velocity of light, fc = carrier frequency and d = distance between transmitter and receiver [2].

For log-distance path loss with shadowing the path loss is denoted by $\overline{L}_p(d)$, which is

$$\overline{L}_p(d) \propto \left(\frac{d}{d_0}\right)^n,\ d \geq d_0$$

or equivalently,

$$\overline{L}_p(d) = \overline{L}_p(d_0) + 10n\log_{10}\left(\frac{d}{d_0}\right) \text{(dB)},\ d \geq d_0$$

where, n = path loss component, d0 = the close-tin reference distance (typically 1 km for macrocells, 100m for microcells), d = distance between transmitter and receiver [2].

III. Materials And Methods

Calculation of the path loss is usually called prediction. Exact prediction is possible only for simpler cases, such as the above-mentioned free space propagation or the flat-earth model. For practical cases the path loss is calculated using a variety of approximations.

Statistical methods (also called stochastic or empirical) are based on fitting curves with analytical expressions that recreate a set of measured data. Among the most commonly used such methods are Okumura Model, Hata Model, and Lee's Model.

In the cities the density of people is high. So the more accurate loss prediction model will be a good help for the BTS mapping for optimum network design. Among the

Radio Propagation Models (RPM) city models are to be analysed in this paper to find the best fitting city model. The well known propagation models for urban areas are:

i) Okumura Model

ii) Hata Model

iii) Lee's Model

A. Okumura Model

The Okumura model for urban areas is a Radio propagation model that was built using the data collected in the city of Tokyo, Japan. This model is applicable for frequencies in the range of 150 MHz to 1920 MHz and distances of 1 km to 100 km. It can be used for the base stations antenna heights ranging from 30m to 100m.

To determine path loss using Okumara's model, the free space path loss between the points of interest is first determined and then the value of $A_{mu}(f,d)$ is added to it along with correction factors according to the type of terrain. The expression of the model [1]:

$$L_{50}(dB) = L_F + A_{mu}(f,d) - G(h_{te}) - G(h_{re}) - G_{AREA} \quad \text{..... (1)}$$

where, L_{50} = The 50th percentile (i.e. median) value of propagation path loss., L_F = The Free Space propagation Loss in dB, A_{mu} = Median attenuation.relative to free space in dB, $G(h_{te})$ = The base station antenna height gain factor, $G(h_{re})$ = Mobile antenna height gain factor and G_{AREA} = The gain due to the type of environment.

$$G(h_{te}) = 20\log_{10}\left(\frac{h_{te}}{200}\right);\quad 100m \rangle h_{te} \rangle 30m$$

$$G(h_{re}) = 10\log_{10}\left(\frac{h_{re}}{3}\right);\quad h_{re} \leq 3m$$

$$G(h_{re}) = 20\log_{10}\left(\frac{h_{re}}{3}\right);\quad 10m \rangle h_{re} \rangle 3m$$

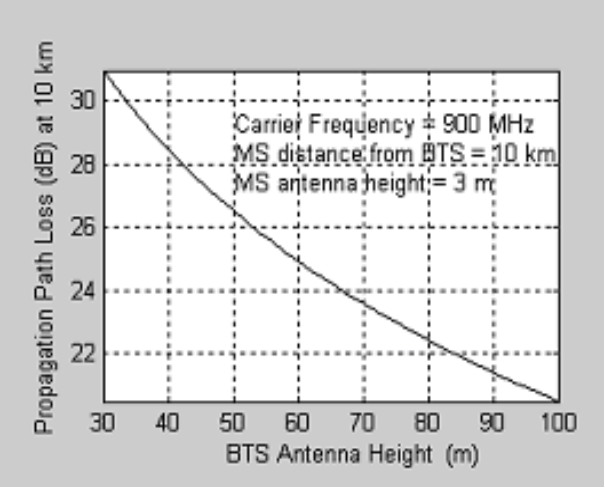


Figure 1. Propagation Path Loss due to the change in BTS antenna height for Okumura model.

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

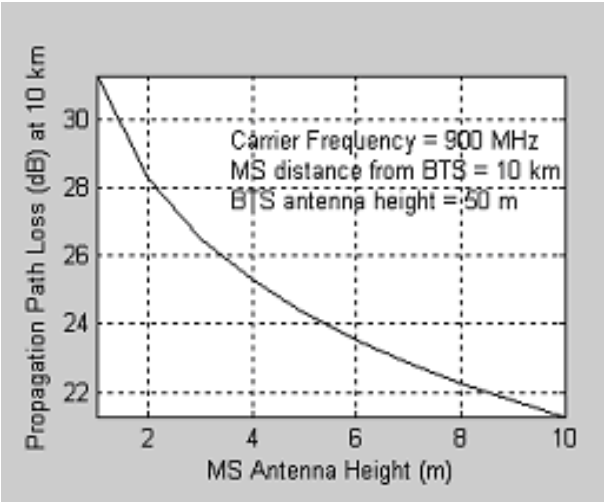


Figure 2. Propagation Path Loss due to the change in MS antenna height for Okumura model.

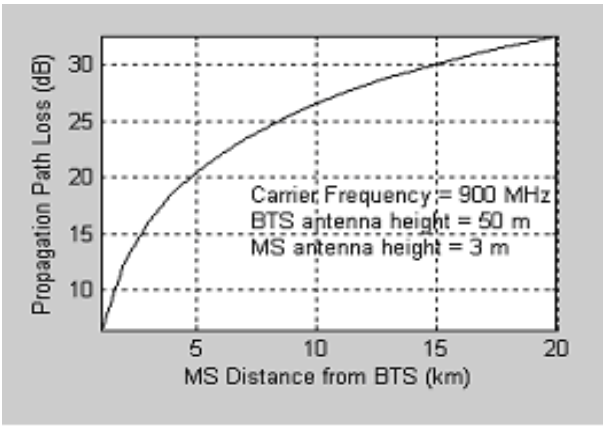


Figure 3. Propagation Path Loss due to T-R separation for Okumura model.

B. Hata Model

Hata Model is based on the Okumara's model where some correction factors are included. It works in the frequencies range from 150 MHz to 1500 MHz.. The standard formula for median path loss in urban areas is given by [1].

$$L_{50}(urban)(dB) = 69.55 + 26.16\log_{10} f_c - 13.82\log_{10} h_{te}) - a(h_{re}) + (44.9 - 6.55\log_{10} h_{te})\log_{10} d \quad \text{.....(2)}$$

where, fc =The frequency from 150 MHz to 1500 MHz, hte = The effective base station antenna height (30m to 200m), hre = The effective mobile antenna height (1m to 10m), d = The trnasmitter-receiver (T-R) distance in km and a(hre) = The correction factor for effective mobile antenna hieght which is a function of the size of coverage area. For a large city it is given by,

$$a(h_{re}) = 8.29(\log_{10} 1.54h_{re})^2 - 1.1dB$$

for $f_c \le 300MH_Z$

$$a(h_{re}) = 3.2(\log_{10} 11.75h_{re})^2 - 4.97dB$$

for $f_c \ge 300MH_Z$

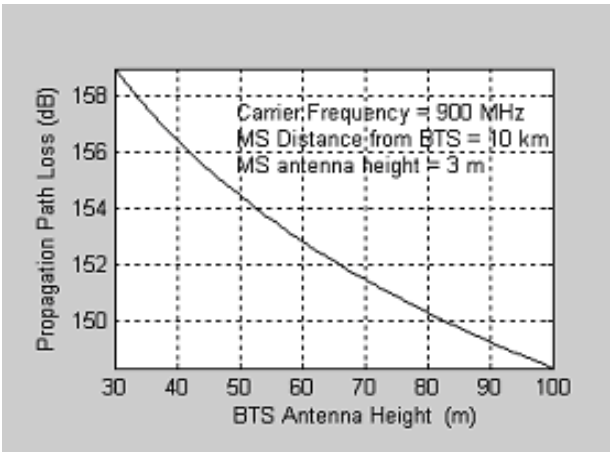


Figure 4. Propagation Path Loss due to the change in BTS antenna height for Hata model.

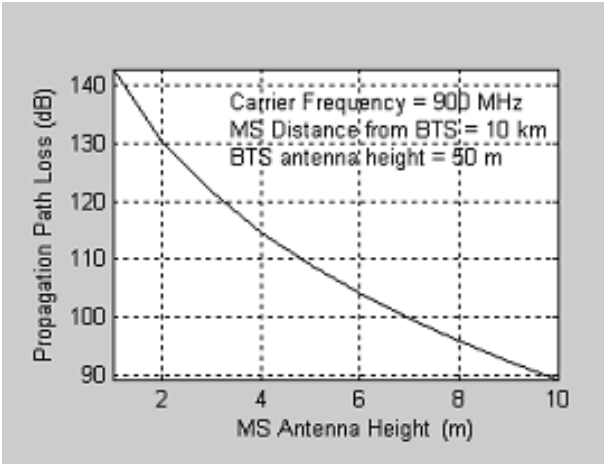


Figure 5. Propagation Path Loss due to the change in MS antenna height for Hata model.

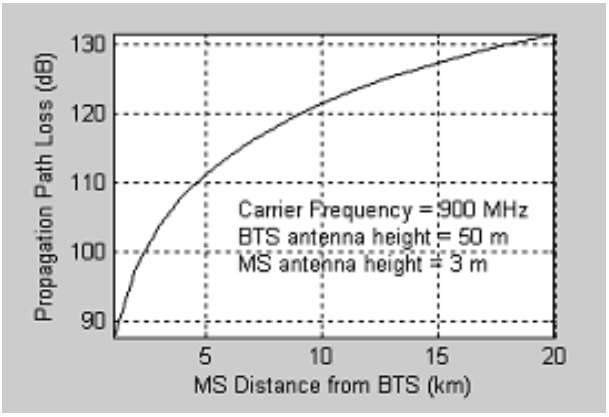


Figure 6. Propagation Path Loss due to T-R separation for Hata model.

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

C. Lee Model

Lee's path loss model is based on empirical data chosen so as to model a flat terrain. Large errors arise when the model is applied to a non-terrain. However, Lee's model has been known to be more of a "North American model" than that of Hata.

The propagation loss calculated as:

$$L(dBm) = 124 + 30.5\log_{10}\left(\frac{d}{d_0}\right) + 10k\log_{10}\left(\frac{f}{f_c}\right) - \alpha_0 \quad \text{.....(3)}$$

where, d is in km, f and fc is in MHz, k = 2 for fc < 450 MHz and in suburban/open area and 3 for fc > 450 MHz and in urban area, d0 = 1.6 km.

In this analysis the parameter values taken for calculations are:

Carrier Frequency fc = 900 MHz

Nominal (Calibration) Distance do = 1.6 km

Base Mobile Station Antenna hb = 30.48 m

Mobile Station Antenna Height hm = 3 m

Base Station Transmit Power Pb = 10 W

Base Station Antenna Gain Gb = 6 dB

Mobile Station Antenna Gain Gm = 0 dB with respect to isotropic antenna.

f is the transmitted frequency, d is the Transmitter-Receiver distance and α0 is a correction factor to account for BS and MS antenna heights, transmit powers and antenna gains that differ from the nominal values. As such, when the prevailing conditions differ from the nominal ones, then α0 is given by:

$$\alpha_0 = 10\log_{10}(\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5)$$

where:

α1 = (new BTS antenna height (m) / 30.48 m)2

α2 = (new MS antenna height (m) / 3 m)v ;for MS antenna height < 3, v =1 and for MS antenna height >3, v =2;

α3 = (new transmitter power / 10 W)2, in this paper the value of α3 is taken 1.

α4 = new BS antenna gain correction factor = (Gb / 4), Here α4 is considered 6 dB.

α5 = frequency correction factor = (f/fc)-n ; for 2<n<3 and f and fc is in MHz. fc is taken 900 MHz.

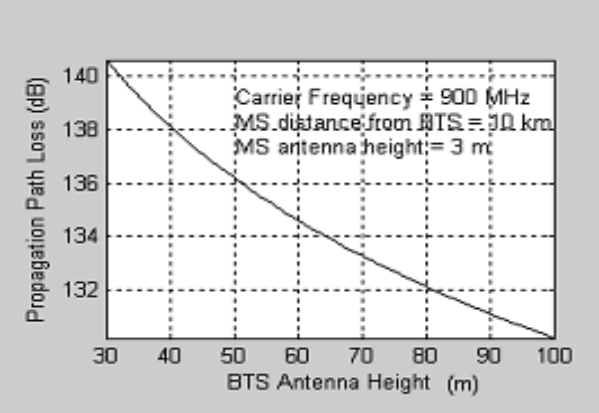


Figure 7. Propagation Path Loss due to the change in BTS antenna height for Lee model.

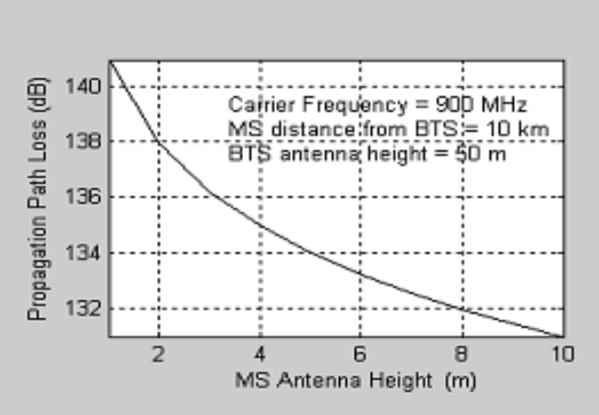


Figure 8. Propagation Path Loss due to the change in MS antenna height for Lee model.

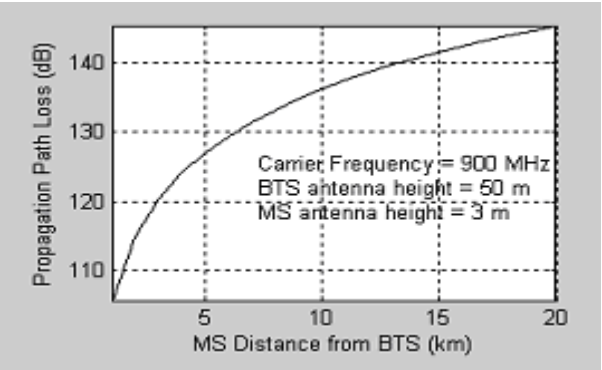


Figure 9. Propagation Path Loss due to the T-R separation for Lee Model

IV. RESULT

From Fig. 10, it is seen that the propagation path loss decreases due to the increase in BTS antenna height for all the models. For Hata model the loss is maximum, for Lee model the loss is medium and for Okumura model the loss is minimum. From Fig. 11, it is seen that the propagation path loss increases with the decrease in MS antenna height for all the models. For Lee model the loss is maximum, for Hata

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

model the loss is medium and for Okumura model the loss is minimum.

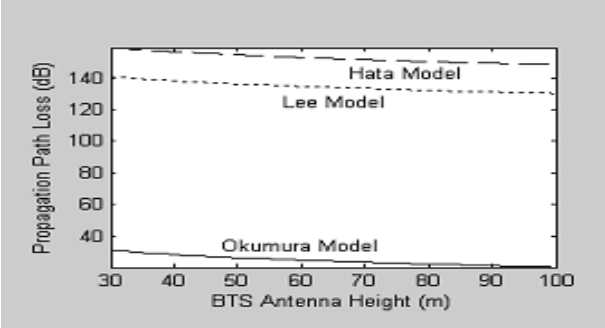


Figure 10. Comparison of Propagation Path Loss due to the change in BTS antenna height.

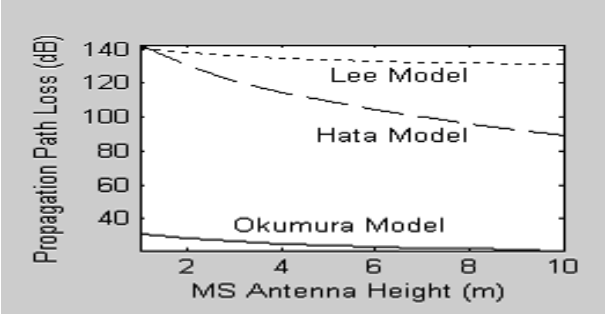


Figure 11. Comparison of Propagation Path Loss due to the change in MS antenna height.

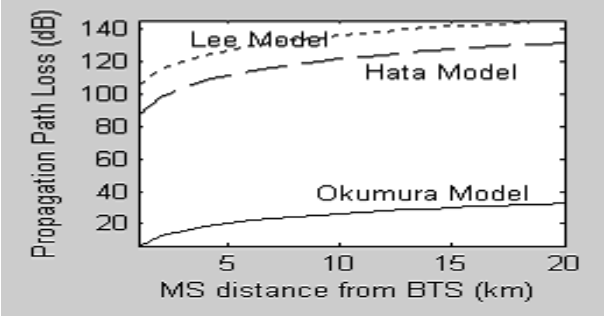


Figure 12. Comparison of Propagation Path Loss due to T-R separation in BTS antenna height.

From Fig. 12 it is seen that for Lee model the propagation path loss is highest due to the increase in MS distance from BTS than the other two models and Okumura model has the lowest path loss. From the analyses it is seen that overall Okumura model shows the better performance than that of the other two models.

V. CONCLUSION

In this paper, three widely known large scale propagation models are studied and analyzed. The analyses and simulation was done to find out the path loss by varying the BTS antenna height, MS antenna height, and the T-R separation. Okumura model was seen to represent low power loss levels in the curves. The result of this analysis will help the network designers to choose the proper model in the field applications. Further up-gradation in this result can be possible for the higher range of carrier frequency.

REFERENCES

[1] Theodore.S.R., 2006, "Wireless Communications Principles and Practice", Prentice-Hall, India.

[2] J. W. Mark, Weihua Zhuang, 2005, "Wireless Communications and Networking", Prentice-Hall, India.

[3] F. D. Alotaibi and A. A. Ali, April 2008, "Tuning of lee path loss model based on recent RF measurements in 400 MHz conducted in Riyadh city, Saudi Arabia," The Arabian Journal for Science and Engineering, Vol 33, no 1B, pp. 145–152.

[4] http:// en.wikipedia.org/wiki/Radio_Propagation_model, June, 2008.

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

Performance Evaluation of TCP over Mobile Ad-hoc Networks

Foez ahmed[1], Sateesh Kumar Pradhan[1], Nayeema Islam[2] , and Sumon Kumar Debnath[3]

[1]Dept. of Computer Networks Engineering, College of Computer Science, King Khalid University, Kingdom of Saudi Arabia
[2]Department of Information & Communication Engineering, Rajshahi University, Rajshahi- 6205, Bangladesh
[3]Dept. of Computer Science and Telecommunication Engineering, Noakhali Science and Technology University, Bangladesh

Abstract— **With the proliferation of mobile computing devices, the demand for continuous network connectivity regardless of physical location has spurred interest in the use of mobile ad hoc networks. Since Transmission Control Protocol (TCP) is the standard network protocol for communication in the internet, any wireless network with Internet service need to be compatible with TCP. TCP is tuned to perform well in traditional wired networks, where packet losses occur mostly because of congestion. However, TCP connections in Ad-hoc mobile networks are plagued by problems such as high bit error rates, frequent route changes, multi-path routing and temporary network partitions. The throughput of TCP over such connection is not satisfactory, because TCP misinterprets the packet loss or delay as congestion and invokes congestion control and avoidance algorithm. In this research, the performance of TCP in Ad-hoc mobile network with high Bit Error rate (BER) and mobility is studied and investigated. Simulation model is implemented and experiments are performed using the Network Simulatior-2 (NS-2).**

Keywords- Ad Hoc Network, TCP, High Bit Error Rate, Route re-computation.

I. INTRODUCTION

An ad hoc network is a collection of mobile nodes forming a temporary wireless network without the aid of any established infrastructure or centralized administration [1]. Since each node in the network is potentially mobile, the topology of an ad-hoc network can be highly dynamic. Ad hoc networks are very useful in emergency search-and-rescue operations, meetings or conventions in which persons wish to quickly share information, and data acquisition operations in inhospitable terrain.

In ad hoc networks all nodes are mobile and can be connected dynamically in an arbitrary manner. All nodes of these networks behave as routers and take part in discovery and maintenance of routes to other nodes in the network. This ad-hoc routing protocols can be divided into two categories: Table-driven routing protocol and On-Demand routing protocols. In table driven routing protocols, consistent and up-to-date routing information of all nodes is maintained at each node. In On-Demand routing protocols, the routes are created when required. When a source wants to send data to a destination, it invokes the route discovery mechanisms to find the path to the destination. In recent years, a variety of new routing protocols targeted specifically at this environment have been developed. There are four multi-hop wireless ad hoc network routing protocols that cover a range of design choices: Destination-Sequenced Distance-Vector (DSDV) [2], Dynamic Source Routing (DSR) [1], On-Demand Distance Vector Routing (AODV) [3] and Temporally Ordered Routing Algorithm (TORA) [4]. While DSDV is a table-driven routing protocol, TORA, DSR, AODV, fall under the On-demand routing protocols category.

As the standard network protocol on the Internet, the use of the TCP/IP over the ad-hoc network is a certainty. It is because that the use of TCP/IP supports a large number of popular applications and allows seamless integration with the internet. In addition, TCP/IP provides network-level consistency for multi-hop networks that consist of hosts using a variety of physical-layer media. However, TCP assumes a relatively reliable underlying network where most packet losses are due to congestion, which conflicts strongly with characteristics of wireless links in ad-hoc mobile network. In mobile ad hoc environment losses are more often caused by high Bit Error Rate (BER), route re-computation, temporary network partition, and multi-path routing [5]**.** If the losses are not due to congestion, then TCP unnecessarily reduces throughput leading to poor performance.

In this paper, the effect of high bit error rate and route re-computation on the performance of TCP in mobile ad hoc network is analyzed. All four mentioned routing protocols are used throughout the experimentations. The TCP variants used in these simulations are TCP Tahoe [6], Reno [7], New Reno [8] and Sack [9].

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

The rest of this paper is organized as follows. In section II the impact of multi-hop wireless networks on TCP performance is discussed briefly. The network topologies together with simulation parameters are presented in section III. Simulation results are also discussed in the same section. Finally, the recommendations for future work in this area and concluding remarks are provided in section IV.

II. The Problems with TCP in Ad Hoc Networks

This section describes the effects of High Bit Error Rate and Route re-computation on the performance of TCP and the possible reasons behind these effects.

A. Effect of High Bit Error Rate (BER)

Bit errors cause packets to get corrupted resulting in lost TCP data segments or acknowledgements. When acknowledgements do not arrive at the TCP sender within a given short amount of time (the retransmit timeout or RTO), the sender retransmits the segment, exponentially backs off its retransmit timer for the next retransmission, and decreases its congestion window. Repeated errors result in a small congestion window at the sender and causes low throughput.

B. Effect of Route Re-computations

When an old route is no longer available, the routing protocol at the sender attempts to find a new route to the destination. It is possible that discovering a new route may take long. As a result, the TCP sender times out, retransmits a packet, and invokes congestion control mechanisms. This is clearly an undesirable behavior because the TCP connection will be very inefficient. If we imagine a network in which route computations are done frequently (due to high node mobility), the TCP connection will never get an opportunity to transmit at the maximum negotiated rate (i.e., the congestion window will always be significantly smaller than the advertised window size from the receiver).

III. Experimental setup, simulation results and discussions

A. Measurement parameter

The performance metrics of the interest in this study is throughput, which is the measure of how soon an end user is able to receive data. It is determined as the ratio of the total data received by the end user and the connection time. A higher throughput will directly impact the user's perception of the quality of service.

B. Simulation Tools

The results in this study are based on simulations using the network simulator (NS2) from Lawrence Berkeley National Laboratory (LBNL) , with extensions from the MONARCH project at Carnegie Mellon [10]. The extensions include a set of mobile ad hoc routing protocols as well as an 802.11 MAC layer and a radio propagation model.

C. Effect of High Bit Error Rate

To investigate the effect of high BER on the TCP performance two different types of topologies are considered: Grid topology and Chain topology. Both topologies are considered to be static.

C.1. Experimental setup for Chain Topology and Result analysis

A 7-node chain or string topology is shown in fig. 1. In a chain topology, TCP packets travel along a chain of intermediate nodes toward the destination.

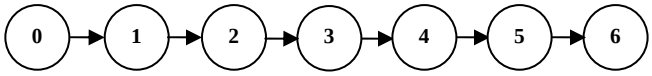


Figure 1. Chain Topology

Simulations are performed to investigate the performance of the TCP flow whose source and destination are placed at both ends of the chain topology. A single wireless channel is shared for transmissions, and only receivers within the transmission range of the sender can receive the packets. The two adjacent nodes are about 200m apart. In this simulation the data rate of the wireless channel is the two-ray ground model with transmission range 250m, carrier sensing range 550m, and interference range 550m. The network model consists of 4, 7 and 10 nodes respectively in a 2200*500 meter flat, rectangular area. At first simulations are performed without introducing any error on the links. Throughputs (in Mbps) of various TCP variants which are TCP Tahoe, TCP Reno, TCP New Reno and Sack over mobile ad hoc networks are identified. For performance comparison all four routing protocols i.e., DSDV, AODV, DSR, TORA are considered. These results for 4-node, 7-node and 10 node networks are shown in table I.

TABLE I. Throughput Of TCP Versions for Chain topology

	4-Node Chain Topology				**7-Node Chain Topology**				**10-Node Chain Topology**			
	DSDV	***AODV***	***DSR***	***TORA***	***DSDV***	***AODV***	***DSR***	***TORA***	***DSDV***	***AODV***	***DSR***	***TORA***
Tahoe	1.248	1.603	1.603	1.602	1.248	1.603	1.603	1.602	1.248	1.603	1.603	1.602
Reno	1.248	1.603	1.603	1.602	1.248	1.603	1.603	1.602	1.248	1.603	1.603	1.602
New Reno	1.313	1.686	1.685	1.685	1.318	1.686	1.686	1.685	1.313	1.686	1.686	1.685
Sack	1.248	1.603	1.603	1.602	1.248	1.603	1.603	1.602	1.248	1.603	1.603	1.602

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

From the table I, it is observed that maximum throughput is achieved for New Reno protocol for all four ad hoc routing protocols. However if routing protocols are compared, AODV shows the best performance. From these results it can be concluded that New Reno-AODV combination shows the best performance for all the sizes of chain topologies. Hence this combination is used to identify the effect of high BER. Simulations are performed to investigate the performance of TCP for a variety of wireless packet error rates ranging from 0.000 to 0.017. Each simulation is performed for 200 seconds.

The observed throughputs for all the networks are plotted in fig. 2. From figure it can be observed that TCP throughput highly degrades with increasing packet error rate that proves its poor performance. Since TCP's congestion control algorithm is responsible for TCP throughput it leads to poor performance in High error rate environment. This figure also compares the impact of varying network sizes in presence of Bit Error. It can be observed that as the network size increased, throughput tends to decrease earlier with increasing PER. For example when the packet error rate is 0.015 the throughput for 4 node chain topology is 0.0296559 Mbps, the throughput of 7 node chain topology is 0.001024 Mbps and finally the throughput of 10 node chain topology is 0.00Mbps. Hence, not only the increasing bit error rates effects the network throughput but also the increasing network size have significant effect on TCP performance.

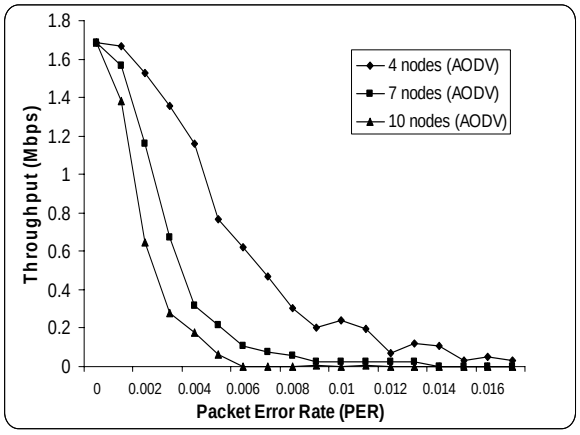


Figure 2. Throughput of TCP New Reno over ad hoc network for PER ranging from 0.0 to 0.016 for various sizes of Chain Topologies.

C.2. Experimental setup for Grid Topology and Result analysis

Three different sizes of Grid topologies are used for simulation. The numbers of nodes of these grid topologies are 16, 25, and 49 respectively. A 16-node grid topology is shown in fig. 3.

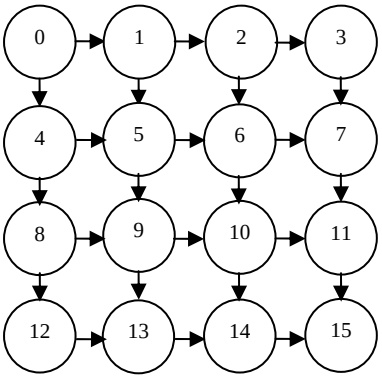


Figure 3. 16-node Grid Topology

The performance of the TCP flow is investigated whose source and destination is placed at two opposite corners of the grid (Bottom-left and Upper-Top of the grid). For the 16-node network the distance between two adjacent nodes is 100 meter. For other two networks the nodes are 50 meters apart. The data rate of the wireless channel is the two-ray ground model with transmission range 250m, carrier sensing range 550m, and interference range 550m. The size of all networks is 500m × 500m. . Again simulations are performed without introducing any error on the links. Throughputs (in Mbps) of all TCP variants all four mentioned routing protocols are calculated. These results are shown in table II.

TABLE II. THROUGHPUT OF TCP VERSIONS FOR GRID TOPOLOGY

	16-Node Grid Topology				**25-Node Grid Topology**				**49-Node Grid Topology**			
	DSDV	***AODV***	***DSR***	***TORA***	***DSDV***	***AODV***	***DSR***	***TORA***	***DSDV***	***AODV***	***DSR***	***TORA***
Tahoe	1.4995	1.6029	1.6024	1.5938	1.2996	1.6029	1.6024	1.5937	1.5677	1.6029	1.7795	1.5929
Reno	1.4981	1.6029	1.6024	1.5945	1.2964	1.6029	1.6024	1.5944	1.5666	1.6029	1.7795	1.5936
New Reno	1.5492	1.6859	1.6853	1.6769	1.3313	1.6859	1.6854	1.6768	1.6310	1.6859	1.7786	1.6760
Sack	1.4980	1.6029	1.6024	1.5938	1.3001	1.6029	1.6024	1.5937	1.5667	1.6029	1.7795	1.5929

For 16-node grid topology New Reno achieves maximum throughput for all the ad hoc routing protocols. AODV routing protocol gives the best performance when New Reno is chosen as the transport layer protocol. Same result is obtained for 25 node grid topology. In case of 49 node grid topology New Reno shows the best performance. However

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

in this case DSR is the dominant routing protocol since it helps to achieve maximum throughput.

Next, errors are generated for all the networks. The packet error rate is varied from 0.000 to 0.017. The calculated throughputs are plotted in fig. 4. From figure 4 can be observed that increasing error decreases the TCP throughput for all network sizes. For 25-node network, throughput decreases rapidly than 16-node network with increasing PER. For 49-node network throughput decreases slowly than other two. Again, the reduction of throughput of TCP is due to mistaking wireless error as an indication of congestion.

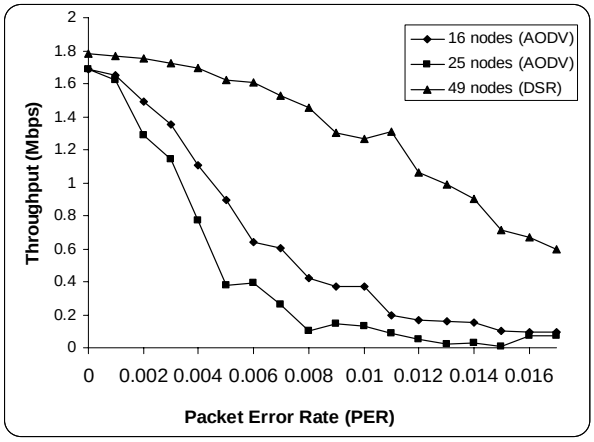


Figure 3. Throughput of TCP New Reno over ad-hoc network for PER ranging from 0.000 to 0.016 for various sizes of grid topologies

D. Effect of Route Re-computation on the Performance of TCP

The effects of route re-computation are investigated for AODV and TORA routing protocols.

D.1. Effect of Route Re-computation for AODV Routing Protocol

To experiment the effect of route re-computation on the performance of TCP the following network topology (fig. 5) is used.

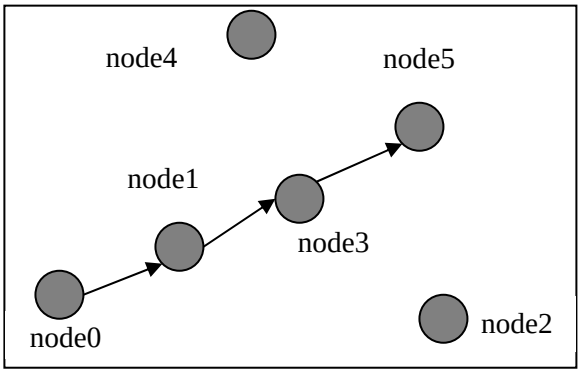


Figure 4. Network Topology before Mobility (AODV Routing Protocol)

There are 6 nodes in this network. The scenario area is 600 m × 600 m and transmission Range of each node is 250 m. All nodes communicate with identical wireless radios, which have a bandwidth of 2Mbps. It is assumed that all links are error free. New Reno is used as the TCP standard. The simulation is run for a period of 200 seconds. The initial positions of the nodes are, node0: (25, 25), node1: (156,143), node2: (396,143), node3: (303,242), node4: (160,331), node5: (399,329).

During simulation node0 starts sending packets at 5 seconds to node5 through the route node0→node1→node3→node5 (figure 5). At time 10 seconds node3 starts moving towards the coordinate position (599,599) at the speed of 30m/s. At the same time node2 also start moving towards the coordination position (303,242) at the same speed as the node3. Finally a new network topology is formed as shown in fig. 6.

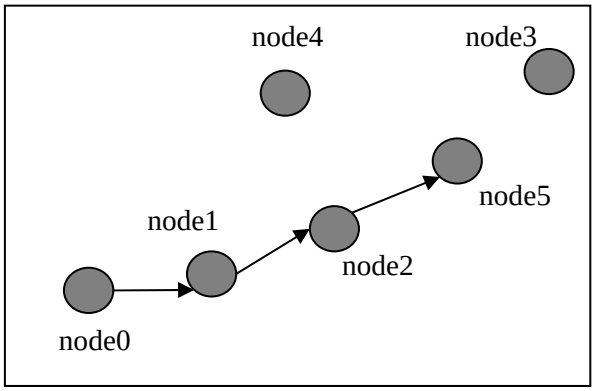


Figure 5. Changed Network Topology after Mobility (AODV Routing Protocol)

Due to mobility node3 moves out of the transmission range and within few seconds the transmission link breaks. Now the old route is no longer available and hence the routing protocol at the sender tries to find a new route to the destination. From the simulation scenario it can be observed that if all other nodes are fixed it is not possible for the sender to find out the new route due the unavailability of the wireless transmission range. To establish a connection a node must move towards the transmission range, which we generate through the movement of node2. At time nearly 40 seconds node2 reaches the transmission range. At the same time a new route is established and node0 again starts sending packets through the new route node0→ node1→node2→node5. It is seen that discovering a new route takes about 35 seconds. Meanwhile the TCP sender has timed out and has invoked congestion control mechanism, which results sever degradation of TCP throughput. Fig. 7 shows the performances of this network for the entire simulation time. From the graph it can be observed that during time interval 10 to 40 seconds no packet is delivered

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

from the source to the destination. Hence the throughput remains zero for that time period. After 40 seconds the connection resumes due to the new route and throughput begins to increase.

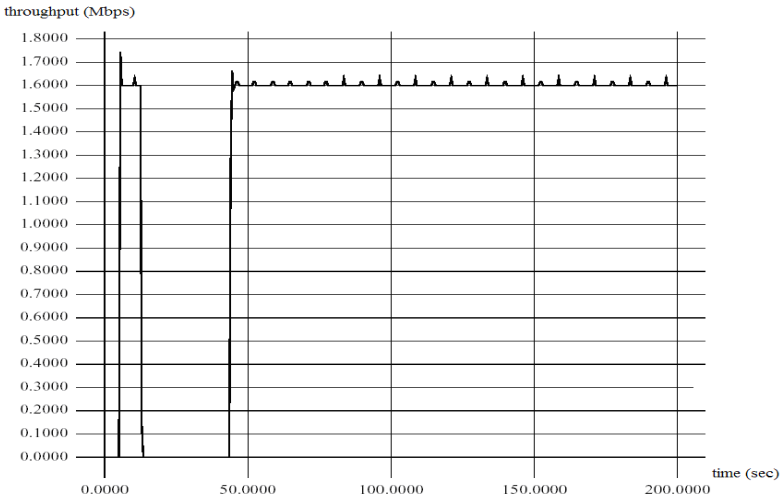


Figure 6. Throughput of TCP over Mobile Ad-hoc network for the entire Simulation time of 200 seconds in presence of Dynamic Network Topology (AODV Routing Protocol).

D.2. Effect of Route Re-computation for TORA Routing Protocol

To investigate the effect of route re-computation on the performance of TCP in case of TORA network protocol the same topology is used as AODV (Figure 5). The network parameters are also same. In this case the sender node0 has established the route through node1 and node2 to send packets to node5. The route is shown in fig. 8

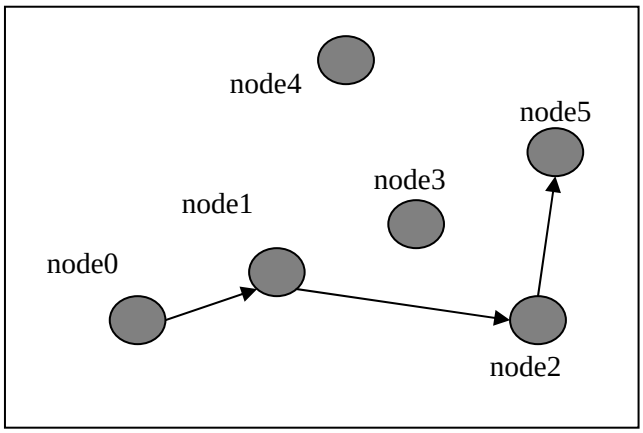


Figure 7. Network Topology before Mobility (TORA Routing Protocol)

At time 10 seconds node2 starts moving towards the coordinate position (599,599) at the speed of 30m/s. When this node moves out of the wireless transmission range the connection breaks and packets start dropping. TCP assumes these packet drops as the indication of congestion and invokes congestion control algorithm. Meanwhile node0 tries to find out a new route and at near about 20 seconds it established a new route through the nodes1 and node3 and again resumes sending packets. The new route is shown in fig. 9.

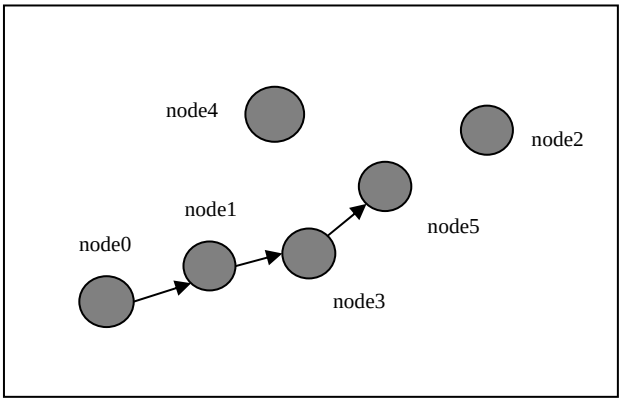


Figure 8. Changed Network Topology after Mobility (TORA Routing Protocol)

The TCP throughput for the entire simulation time is shown in the graph of figure 10. From the graph it is clear that between the time intervals 10 to 20 seconds no packet is successfully delivered to the destination and hence the throughput reduces zero. After 20 seconds when the connection is re-established throughput starts increasing and remains almost same for the remaining time.

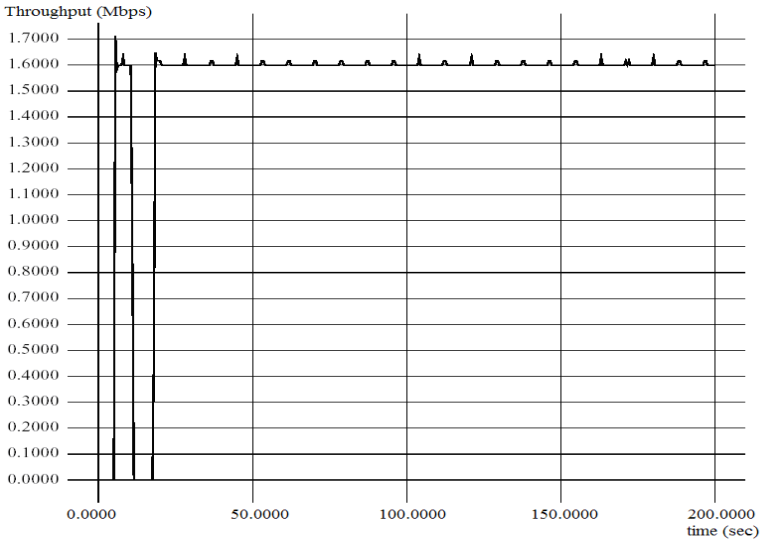


Figure 9. Throughput of TCP over Mobile Ad-hoc network for the entire Simulation time of 200 seconds in presence of Dynamics Network Topology (TORA Routing Protocol)

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

IV. CONCLUSIONS AND RECOMMENDATIONS

In this paper, TCP performance is evaluated in a multi-hop wireless network environment. The impact of high bit error rate of wireless links on the performance of TCP is investigated through simulation for both the chain and grid topologies. TCP cannot differentiate packet lost due to congestion and packet lost due to bit error. For both cases, the TCP sender assumes that networks congestion has occurred and invokes congestion control algorithm. From simulations it is observed that this algorithm results in significant reduction in throughput and unacceptable inter-reactive delays for active connections, thus severely degrading performance. Due to dynamic topology of ad hoc network the route between a sender and receiver changes rapidly. From simulation it is seen that the route change results in link disconnections, which reduces TCP throughput. Also sometimes it takes long to find a new route to resume data transfer. Meanwhile TCP sender times out and invokes congestion control mechanisms. From experiments it is also observed that due to difference of routing protocols, same topology uses different routes to transmit data.

Further investigation can be carried out to study the effects of network partitions and multi-path routing on the TCP performance. The impact that the link-layer has on TCP performance, such as aggregate delay caused by local retransmissions over multiple wireless hops can be experimented through simulation. Experiments can be carried out to evaluate the performance in a congested ad hoc network.

REFERENCES

[1] Johnson, D. and Maltz, D. (1996). "Dynamic source routing in ad hoc wireless networks," in Mobile Computing (ed. T. Imielinski and H. Korth), Kluwer Academic Publishers, Dordrecht, The Netherlands.

[2] Perkins, C., E., and Bhagwat, P., (1994), "Highly dynamic Destination-Sequenced Distance-Vector routing (DSDV) for mobile computers," Proc. of the DIGCOMM '94 Conference on Communications Architectures, Protocols and Applications, pages 234-244, August-1994.

[3] Perkins, C., Elizabeth, M., R., and Samir, R. D., (1999), "Ad hoc on demand distance vector (AODV) routing," Internet draft, IETF, January 1999.

[4] Park, V., and Corson, S., "Temporally Ordered Routing Algorithm (TORA)," Version 1 Internet Draft, draft-ietf-manet-tora-spec- 03.txt, work in progress, June 2001

[5] Dong, X., Lee, D., and Wang, j., "Improving TCP Performance over Ad-Hoc Mobile Networks," Course Project Repost for EE228A, November 2000.

[6] Jacobson, V., "Congestion avoidance and control," Proceedings of the ACM Symposium on Communications Architectures and Protocols, Vol. 18, No. 4, pp. 314-329, Stanford, CA, USA, August 16-18, 1988.

[7] Jacobson, V., "Modified TCP Congestion Avoidance Algorithm," end2end-interest mailing list, April 30, 1990. <URL: ftp://ftp.isi.edu/end2end/end2end-interest-1990.mail.>

[8] Hoe, J. C., "Improving the Start-behavior of a Congestion Control Scheme for TCP," Annual conference of the Association for Computing Machinery's Special Interest Group on Data Communication (ACM SIGCOMM '96), pp.270-280, California, USA, August 26-27, 1996.

[9] Mathis, M., Mahdavi, J., Floyd, S. and Romanow, A., "TCP Selective acknowledgement options," IEFT, RFC 2018 (Status Proposed Standard), 1996. <URL: www.rfc-editor.org/rfc/rfc2018.txt>

[10] Fall., K., and Varadhan., K.,"NS-Documentation, http://www.isi.edu/nsnam/ns/ns- documentation.html," December 2006.

AUTHORS PROFILE

Foez Ahmed joined as a Lecturer in the Dept. of Electronics and Communication Engineering, Northern University Bangladesh (NUB), Dhaka, Bangladesh in the 2008. At present he is on leave from Northern University and working as a Lecturer with the Dept of Networks and Communication Engineering, College of Computer Science, King Khalid University, Kingdom of Saudi Arabia. He did his under graduation and post graduation in Information and Communication Engineering in 2007 and 2009 respectively from Rajshahi University, Bangladesh. He has received various Awards and Scholarships for the under graduation and post graduation results. His research interests include Mobile Ad-hoc Networks and Routings, Cognitive Radio Networks, Cooperative Communications, Sensor Networks and Sparse Signal Processing in Wireless Communication.

Sateesh Kumar Pradhan obtained his PhD degree in Computer Science from Berhampur University, India during the year 1999. He joined Berhampur University, as Assistant Professor in the 1987 and promoted to Associate Professor in 1999. He was Head of the Department of Computer Science, Utkal University, India during 2001-2003. He was the Secretary and Vice-President Orissa IT Society from 2003-2005 and 2005-2007 respectively and was the organizing Chair of the International Conference on Information Technology – 2005 (ICIT – 2005). At present he is on leave from Utkal University and working with the

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

Department of Computer Engineering, King Khalid University, KSA. His research interests include Neuron based Algorithms, Computer Architecture, Ad-hoc Network and Computer Forensic.

Nayeema Islam was born 1976 in Bangladesh. She received her M.S. in Telecommunication Engineering from Asian Institute of Technology, Thailand in 2004. Also she received her M.Sc. and B.Sc. in Computer Science and Technology from Rajshahi University, Bangladesh in 1998 and 1997 respectively. She is presently working as Assistant Professor in the department of Information and Communication Engineering, University of Rajshahi since 2004. Her fields of interest include Telecommunications, computer networking, mobile ad-hoc networks and QoS routing.

Sumon Kumar Debnath obtained his M.Sc degree in Information and Communication Engineering from Rajshahi University, Bangladesh during the year 2009. He has joined Noakhali Science and Technology University, Bangladesh and working as a lecturer in the department of Computer Science and Telecommunication Engineering. He is also working as a Assistant Proctor in that university. His research interests include Mobile Ad-hoc network, Sensor Network, MIMO, OFDM, and MCCDMA.

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

Vision Based Game Development Using Human Computer Interaction

Ms.S.Sumathi
Bharath University
Chennai,India

Dr.S.K.Srivatsa
St.Joseph College of Engineering
Chennai,India

Dr.M.Uma Maheswari
Bharath University
Chennai,India

***Abstract*— A Human Computer Interface (HCI) System for playing games is designed here for more natural communication with the machines. The system presented here is a vision-based system for detection of long voluntary eye blinks and interpretation of blink patterns for communication between man and machine. This system replaces the mouse with the human face as a new way to interact with the computer. Facial features (nose tip and eyes) are detected and tracked in real-time to use their actions as mouse events. The coordinates and movement of the nose tip in the live video feed are translated to become the coordinates and movement of the mouse pointer on the application. The left/right eye blinks fire left/right mouse click events. The system works with inexpensive USB cameras and runs at a frame rate of 30 frames per second.**

Keywords: Human Computer Interface (HCI), SSR Filter, Hough transform

I. INTRODUCTION

One of the promising fields in artificial intelligence is HCI. Human-Computer Interface (HCI) can be described as the point of communication between the human and a computer. HCI aims to use human features to interact with the computer. The system tracks the computer user's movements with a video camera and translates them into the movements of the mouse pointer on the screen. The tip of the user's nose can be tracked and captured with a webcam and monitor its movements in order to translate it to some events that communicate with the computer. In our system, the nose tip is the pointing device, because of the location and shape of the nose, as it is located in the middle of the face it is more comfortable to use it as the feature that moves the mouse pointer and defines its coordinates, and it is located on the axis that the face rotates about, so it basically does not change its distinctive convex shape which makes it easier to track as the face moves. We use the human feature Eyes to simulate mouse clicks, so the user can fire their events as he blinks. While different devices were used in HCI (e.g. infrared cameras, sensors, microphones) we used a webcam that affords a moderate resolution and frame rate as the capturing device in order to make the ability of using the program affordable for all individuals.

We present an algorithm that distinguishes true eye blinks from involuntary ones, detects and tracks the desired facial features precisely, and fast enough to be applied in real-time. In our work we are trying to implement this in playing video games. Getting the player physically involved in the game provides a more immersive experience and a feeling of taking direct part rather than just playing as an external beholder. Already motion sensors have been implemented to recognize physical activity, we can even use finger gestures in playing games but we increase human computer interaction by using our eyes in playing games.

We propose an accurate algorithm that distinguishes true eye blinks from involuntary ones, detects and measures their duration, and fast enough to be applied in real-time to control a non – intrusive interface for computer users in playing games. This system can be used in playing games like first shooter game, a role playing game and an action game. When we compare eye-based and mouse-based control it is found that using an eye

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

tracker can increase the immersion and leads to a stronger feeling of being part of the game. We are tracking the nose movements and eye blinks for playing the games, the nose movements are interfaced with the mouse movements and the eye blinks are interfaced with the mouse clicks. The eye-movement or eye blink controlled human-computer interface systems are very useful for persons who cannot speak or use hands to communicate. There are no lighting requirements or offline templates needed for the proper functioning of the system. It works with inexpensive USB cameras and runs at a frame rate of 30 frames per second.

The automatic initialization phase is triggered by the analysis of the involuntary blinking of the current computer user, which creates an online template of the eye to be used for tracking. This phase occurs each time the current correlation score of the tracked eye falls below a defined threshold in order to allow the system to recover and regain its accuracy in detecting the blinks. This system can be utilized by users for applications that require mouse clicks as input for e.g. games. The main contribution of this paper is to provide a reimplementation of this system that is able to run in real time at 30 frames per second on readily available and affordable webcams in playing video games.

II. RELATED WORK

With the growth of attention about computer vision, the interest in HCI has increased proportionally. Different human features and monitoring devices were used to achieve HCI, but during our research we were only into works that involved the use of facial features and webcams.

The current evolution of computer technologies has enhanced various applications in human-computer interface. Face and gesture recognition is a part of this field, which can be applied in various applications such as in robotic, security system, drivers monitor, and video coding system.

We noticed a large diversity of the facial features that were selected, the way they were detected and tracked, and the functionality that they presented for the HCI. Researchers chose different facial features: eye pupils, eyebrows, nose tip, lips, eye lids' corners, mouth corners for each of which they provided an explanation to choose that particular one.

Different detection techniques were applied where the goal was to achieve more accurate results with less processing time. To control the mouse pointer various points were tracked ranging from the middle distance between the eyes, the middle distance between the eyebrows, to the nose tip. To simulate mouse clicks; eye blinks, mouth opening/closing, and sometimes eyebrow movement were used. Each HCI method that we read about had some drawbacks, some methods used expensive equipments, some were not fast enough to achieve real-time execution, and others were not robust and precise enough to replace the mouse. We tried to profit from the experience that other researchers gained in the HCI field and added our own ideas to produce an application that is fast, robust, and useable.

Eye movement events detected in EOG signals such as saccades, fixations and blinks have been used to control robots or a wearable system for medical care givers. Patmore et al. described a system that provides a pointing device for people with physical disabilities. All of these systems use basic eye movements or eye-gaze direction but they do not implement movement sequences which provide a more versatile input modality for gaming applications.

III. SYSTEM OVERVIEW

This system design can be broken down into three modules, (1) Facial Features tracking (2) Integrating Nose tip movements with the mouse cursor (3) Replacing the eye blinks with the mouse click events.

Face Detection

In this module, we propose a real-time face detection algorithm using Six-Segmented Rectangular (SSR) filter, distance information, and template matching technique. Since human face is a dynamic object and has a high degree of variability,

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

we propose the method combine both feature-based and image-based approach to detect the point between the eyes by using Six-Segmented Rectangular filter (SSR filter). The proposed SSR filter, which is the rectangle divided into 6 segments, operates by using the concept of bright-dark relation around Between-the-Eyes area. Between-the-Eyes is selected as face representative in our detection because its characteristic is common to most people and is easily seen for a wide range of face orientation.

Firstly, we scan a certain size of rectangle divided into six segments throughout the face image. Then their bright-dark relations are tested if its center can be a candidate of Between-the-Eyes. Next, the distance information obtained from stereo camera and template matching is applied to detect the true Between-the-Eyes among candidates. Between-the-Eyes has dark part (eyes and eyebrows) on both sides, and has comparably bright part on upper side (forehead), and lower side (nose and cheekbone). This characteristic is stable for any facial expression.

We use an intermediate representation of image called integral image to calculate sums of pixel values in each segment of SSR filter. Firstly, SSR filter is scanned on the image and the average gray level of each segment is calculated from integral image. Then, the bright-dark relations between each segment are tested to see whether its center can be a candidate point for Between-the- Eyes. Next, the stereo camera is used to find the distance information and the suitable Between-the- Eyes template size.. Finally the true Between-the-Eyes can be detected.

1) Integral Image

The SSR filter is computed by using intermediate representation for image called integral image. For the original image i(x, y), the integral image is defined as,

$$ii(x,y) = \sum_{x' \le x} \sum_{y' \le y} i(x', y') \qquad (1)$$

The integral image can be computed in one pass over the original image by the following pair of recurrences.

$$s(x, y) = s(x, y - 1) + i(x, y) \qquad (2)$$

$$ii(x, y) = ii(x - 1, y) + s(x, y) \qquad (3)$$

Where $s(x, y)$ is the cumulative row sum,

$s(x, -1) = 0$, and $ii(-1, y) = 0$.

Using the integral image, the sum of pixels within rectangle D (r_s) can be computed at high speed with four array references as shown in Fig.1.

$$s_r = (ii(x, y) + ii(x - W, y - L)) - (ii(x - W, y) + ii(x, y - L)) \qquad (4)$$

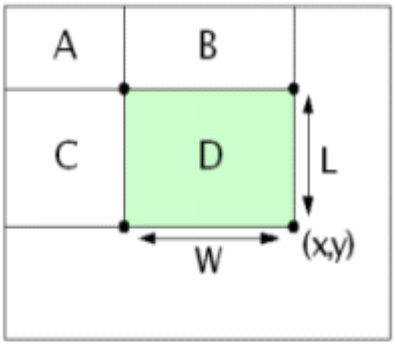


Figure 1. Integral Image

2) SSR filter

At the beginning, a rectangle is scanned throughout the input image. This rectangle is segmented into six segments as shown in Fig.2 (a).

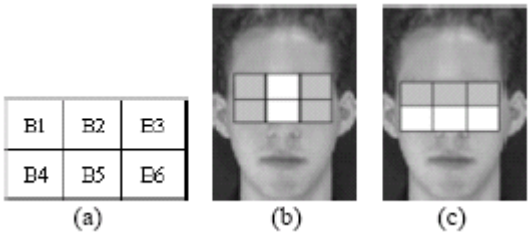


Figure 2. SSR Filter

We denote the total sum of pixel value of each segment (B1 B6) as 1 6 b b S S. The proposed SSR filter is used to detect the Between-the-Eyes based on two characteristics of face geometry.

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

(1) The nose area (n S) is brighter than the right and left eye area (er S and el S , respectively) as shown in Fig.2 (b), where

$$S_n = S_{b2} + S_{b5}$$
$$S_{er} = S_{b1} + S_{b4}$$
$$S_{el} = S_{b3} + S_{b6}$$

Then,

$$S_n > S_{er} \quad (5)$$
$$S_n > S_{el} \quad (6)$$

(2) The eye area (both eyes and eyebrows) (e S) is relatively darker than the cheekbone area (including nose) (c S) as shown in Fig. 2 (c), where

$$S_e = S_{b1} + S_{b2} + S_{b3}$$
$$S_c = S_{b4} + S_{b5} + S_{b6}$$

Then,

$$S_e < S_c \quad (7)$$

When expression (5), (6), and (7) are all satisfied, the center of the rectangle can be a candidate for Between-the-Eyes.

Figure 3. Between-the-Eyes candidates from SSR filter

In Fig.3 (b), the Between-the-Eyes candidate area is displayed as the white areas and the non-candidate area is displayed as the black part. By performing labeling process on Fig. 3 (b), the result of using SSR filter to detect Between-the-Eyes candidates is shown.

Because the SSR filter extracts not only the true Between-the- Eyes but also some false candidates, so we use the average Between-the-Eyes template matching technique to solve this problem.

To evaluate the candidates, we define the Between the- Eyes pattern as p_{mn} (m=0,...,31, n = 0, ..., 15) . Then right and left half of p_{mn} is re-defined again separately as p^r_{ij} (i=0,...,15, j = 3, ..., 15) and p^1_{ij} (i=0,...,15, j = 3, ..., 15), respectively, each has been converted to have average value of 128 and standard deviation of 64.

Then the left mismatching value (D_l) and the right mismatching value (D_r) are calculated by using the following equation.

$$D_l = \sum \frac{(p_{ij^l} - t_{ij^l})^2}{v_{ij^l}} \quad (8)$$

$$D_r = \sum \frac{(p_{ij^r} - t_{ij^r})^2}{v_{ij^r}} \quad (9)$$

The processing flow of Real-Time face detection system is shown in Fig. 4.

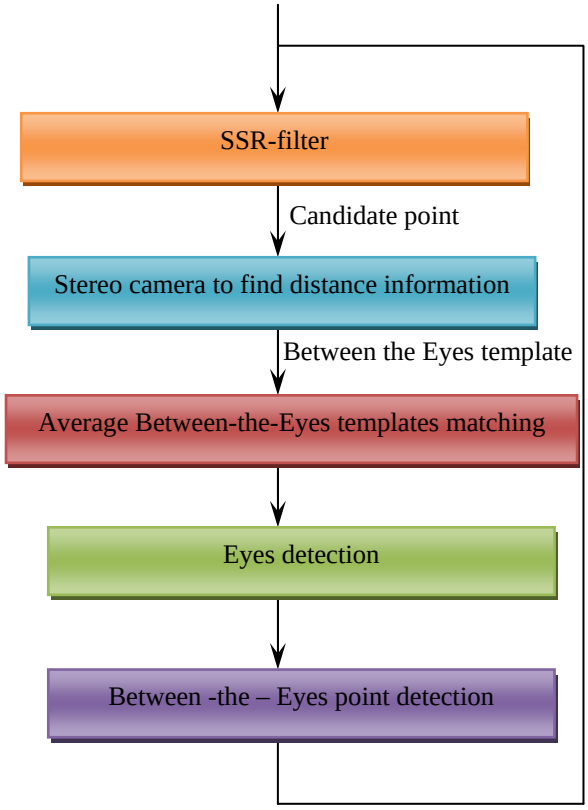


Figure 4. Processing Flow of Real-Time Face Detection

After extracting the templates, we pass them to the support vector machine in order to classify them. Positive classification results mean true faces, while

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

negative ones mean false faces. Since the program will be used by one person at a time, we need to pick one of the positive results as the final detected face. To achieve that, we pick the highest positive result, but before doing so, we will multiply each positive result by the area of the cluster that its template represents.

Finding the Nose Tip

After locating the eyes, the final step is to find the nose tip. From figure 5 we can see that the blue line defines a perfect square of the pupils and outside corners of the mouth; the nose tip should fall inside this square, so this square becomes our region of interest in finding the nose tip.

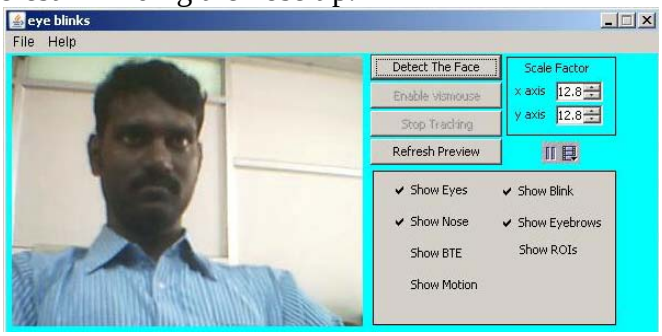


Figure 5 The square that forms the ROI

The nose tip has a convex shape so it collects more light than other features in the ROI because it is closer to the light source. In horizontal intensity profiles we add vertically to each line the values of the lines that precedes it in the ROI, so since that the nose bridge is brighter than the surrounding features the values should accumulate faster at the bridge location; in other words the maximum value of the horizontal profile gives us the 'x' coordinate of the nose tip.

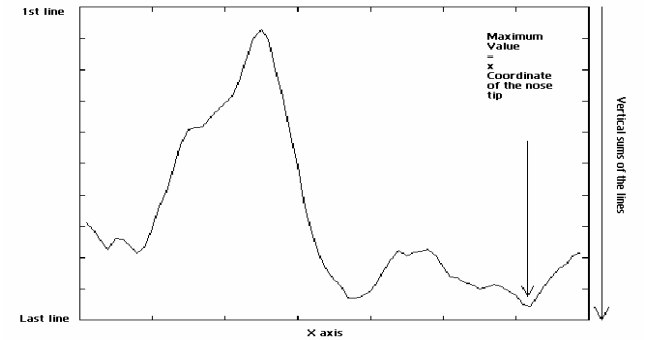


Figure 6:The horizontal profile of ROI

In vertical intensity profiles we add horizontally to each column the values of the columns that precedes it in the ROI the same as in the horizontal profile, the values accumulate faster at the nose tip position so the maximum value gives us the 'y' coordinate of the nose tip. From both, the horizontal and vertical profiles we were able to locate the nose tip position.

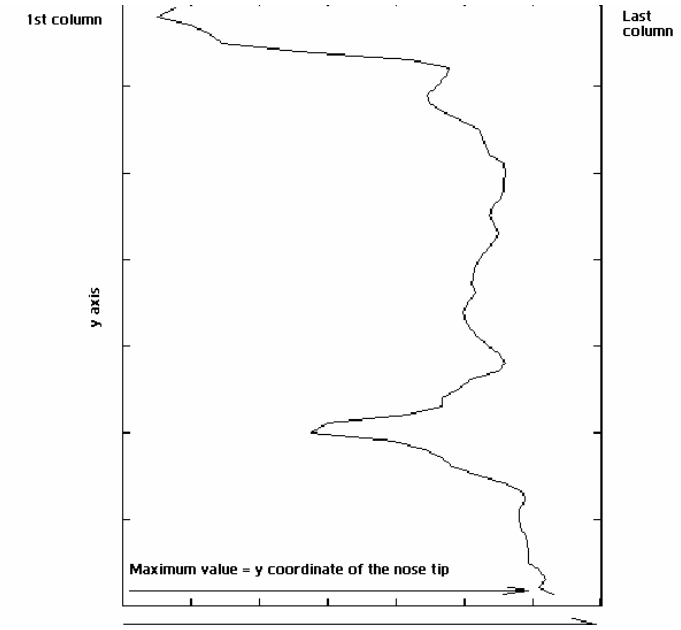


Figure 7: Horizontal Sums of the columns

After locating the nose bridge we need to find the nose tip on that bridge. Since each NBP represents the brightest S2 sector on the line it belongs to, and that S2 sector contains the accumulated vertical sum of the intensities in that sector from the first line to the line it belongs to, we will be using this information to locate the nose tip. Nose trills are dark areas, and the portion that they add to the accumulated sum in the horizontal profile is smaller than the contribution of other areas; in other words each NBP will add with its S2 sector a certain amount to the accumulated sum in the horizontal profile, but the NBP at the nose trills location will add a smaller amount, we will notice a local minima at the nose trills location, by locating this local minima we take the NBP that corresponds to it as the nose trills location, and the next step is to look for the nose tip above the nose trills. Since the nose tip is brighter than other features it will donate with its S2 sector to the accumulated sum more than

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

other NBPs, which means a local maxima in the first derivate, so the location of the nose tip is the location of the NBP that corresponds to the local maxima that is above the local minima in the first derivate. Tracking the nose tip will be achieved by template matching inside the ROI.

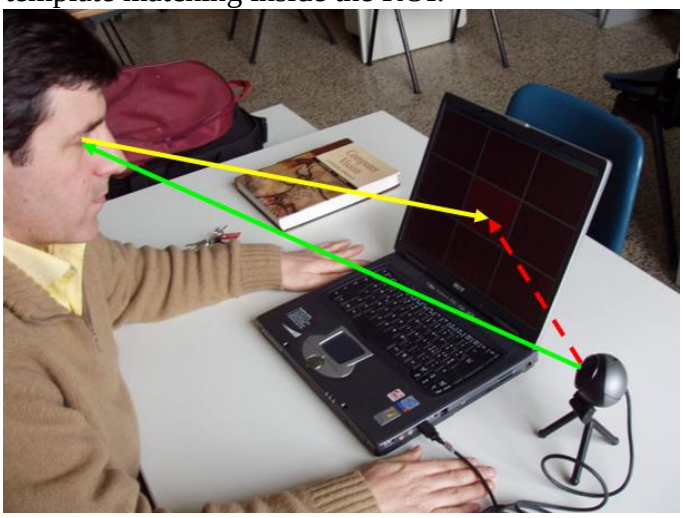


Figure 8:The final result of the face and eye detection process

Hough Transform

The Hough transform is a technique which can be used to isolate features of a particular shape within an image. In our proposal we use this to find our eye brows. The classical Hough transform is most commonly used for the detection of regular curves such as lines, circles, ellipses, etc.

The Hough transform can be used to identify the parameter of a curve which best fits a set of given edge points. This edge description is commonly obtained from a feature detecting operator such as the Roberts Cross, Sobel or Canny edge detector and may be noisy, i.e. it may contain multiple edge fragments corresponding to a single whole feature. Furthermore, as the output of an edge detector defines only where features are in an image, the work of the Hough transform is to determine both what the features are (i.e. to detect the feature(s) for which it has a parametric (or other) description) and how many of them exist in the image. To find the eyebrow line from the set of thresholding points we apply the Hough transform. Sometimes Hough transform gives several lines so we approximate them to a final line which is the eye brow.

Motion Detection

To detect motion in a certain region we subtract the pixels in that region from the same pixels of the previous frame, and at a given location (x,y); if the absolute value of the subtraction was larger than a certain threshold, we consider a motion at that pixel.

Blink Detection

We apply blink detection in the eye's ROI before finding the eye's new exact location. The blink detection process is run only if the eye is not moving, because when a person uses the mouse and wants to click, he moves the pointer to the desired location, stops, and then clicks, so basically the same for using the face, the user moves the pointer with the tip of the nose, stops, then blinks. To detect a blink we apply motion detection in the eye's ROI; if the number of motion pixels in the ROI is larger than a certain threshold we consider that a blink was detected, because if the eye is still, and we are detecting a motion in the eye's ROI, that means that the eyelid is moving which means a blink. In order to avoid multiple blinks detection while they are a single blink (because motion pixels will appear while the eye is closing and reopening), the user can set the blink's length, so all blinks which are detected in the period of the first detected blink are omitted.

IV Conclusion

The proposed system is the best system for the users to play games interactively. The automatic initialization phase is greatly simplified in this system, with no loss of accuracy in locating the user's eyes and choosing a suitable open eye template. Another improvement in this system is, it is compatible with inexpensive USB cameras, as opposed to the high- resolution cameras. The experiments indicate that the system performs equally well in extreme lighting conditions. The accuracy percentages in all the cases were approximately the same as those that were retrieved in normal lighting conditions.

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

Another important consideration is the placement and orientation of camera. The experiments showed that placing the camera below the user's head resulted in desirable functioning of the system. However, if the camera is placed too high above the user's head, in such a way that it is aiming down at the user at a significant angle, the blink detection is no longer as accurate. This is caused by the very small amount of variation in correlation scores as the user blinks, since nearly all that is visible to the camera is the eyelid of the user. Thus, when positioning the camera, it is beneficial to the detection accuracy to maximize the degree of variation between the open and closed eye images of the user.Higher frame rates and finer camera resolutions could lead to more robust eye detection that is less restrictive on the user, while increased processing power could be used to enhance the tracking algorithm to more accurately follow the user's eye and recover more gracefully when it is lost.

REFERENCES

1.M. Bartlett, G. Littlewort, B. Braathen, T. Sejnowski, and J. Movellan. A prototype for automatic recognition of Spontaneous facial actions. In S. Becker, S. Thrun, and K. Obermayer, editors, Advances in Neural Information Processing Systems, volume 15, pages 1271–1278, 2003.

2. T. N. Bhaskar, F. T. Keat, S. Ranganath, and Y.V. Venkatesh.Blink detection and eye tracking for eye localization. In Proc. TENCON 2003, volume 2, pages 821–824, 2003.

3. A. Haro, M. Flickner, and I. Essa. Detecting and tracking eyes by using their physiological properties, dynamics, and appearance. In IEEE International Conference on Computer Vision and Pattern Recognition, volume 1, pages 163–168,2000.

4.Q. Ji, H. Wechsler, A. Duchowski, and M. Flickner. Editorial: special issue: eye detection and tracking. Comput.Vis.Image Underst., 98(1):1–3, 2005.

5.D. G. Lowe. Distinctive image features from scale-invariant keypoints. Int. J. of Computer Vision, 60(2):91–110, 2004.

6. M. Valstar, M. Pantic, Z. Ambadar, and J. Cohn. Spontaneous vs. posed facial behavior: Automatic analysis of brow actions. In Int. Conf. on Multimodal Interfaces, pages 162–170, 2006.

AUTHORS PROFILE

She has completed her Bachelor of engineering in Computer Science from Raja Rajeswari engineering College,Anna University and Master of Engineering from P.S.N.A Engineering college ,Anna University.She is currently doing her Ph.D in Bharath University.Her research interest includes image processing,computer vision and database management systems.

Dr.S.K.Srivatsa was born at Bangalore on 21st July 1945.He received his Bachelor of Electronics and Telecommunication Engineering degree (Honors) from Jadavpur University(securing first rank and two medals).Master degree in Electrical Commuication Engineering(With distinction)from Indian Institute of Science and Ph.D also from Indian Institute of Scienc,Bangalore.In July 2005,he retired as professor of Electronics engineering from Anna University.He has taught twenty-two different Courses at P.G level during the last 33 years.He has functioned as a Member of the Board of studies in some Educational Institutions.His name is included in the Computer Society of India database of Resource professionals.He has received about a dozen awards.He has produced 24 Ph.Ds.He is the author of well over 400 publications.

.

Dr.M.Uma maheswari received her bachelor of science in Computer Science from Bharathidasan university in 1995, Master of Computer Applications in Computer Sience from Bharathidasan University in1998,M.Phil in Computer Science from Alagappa University,Karaikudi in 2005,Master of Technology in Computer Science from Mahatma Gandhi Kasi Vidyapeeth university in 2005 and Ph.D in Computer Science from Magadh Universty,Bodh Gaya in 2007.She has 10 years of teahing experience and guided 150 M.C.A projects,23 M.Tech projects and 6 Ph.D research works.

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

Path Traversal Penalty in File Systems

M.I. Lali [*1], F. Ahsan [*2], A.F.M. Ishaq [*#3]

[*] *Department of Computer Science, COMSATS Institute of Information Technology Islamabad, Pakistan*

[1] mikramullah@comsats.edu.pk

[2] fahsan@comsats.edu.pk

[#] *Present address*

SZABIST, Dubai International Academic City, Dubai, UAE

[3] faiz@szabist.ac.ae

***Abstract*—File systems are used to manage data in the form of files and directories. These directories are hierarchical in nature. Access to the stored data is achieved by traversing through the path from root level to the respective directory containing the required file. This complex nature of data storage mechanism has significant effects on the performance of file systems in terms of accessibility. For considering new optimizations for file system design, it is important to study existing ones. Therefore, we designed a benchmark application to measure the penalty over path traversal in different file systems. Here, we present our results for the impact of directory depth in Windows FAT32, NTFS, Linux EXT-2 and Solaris UFS files systems. Overall, It is found that there is a considerable performance degradation as we go deeper along the directory levels in all these file systems.**

***Index Terms*—File System Benchmark, File Server, File Systems, File Access Efficiency, Directory Depth.**

I. INTRODUCTION

Computers are used for accessing and retrieving information in the form of data. The data is stored on storage media and managed by file systems which have become an indispensable part of modern operating systems. Consistency and efficiency of file systems affects the reliability and performance of most of the running applications [1]. Thus, a prime concern of the researchers is to develop efficient and reliable file systems. To achieve the objective of efficiency and reliability, there has been always a need to explore different new possibilities in the area. These new possibilities can only be found by thorough examination of the existing file systems. The prominent Benchmarks to find the effect of different parameters over efficiency of file systems are shown in [2]. File system efficiency is greatly dependent on the data layout which depends on the structure of the file system [3]. Furthermore, the paper says that in a specific data layout, file access time depends on the directory depth at which it is located. File access efficiency can be measured and compared for different levels of directory depth by benchmarking applications.

Most of the existing file system benchmarks measure the file read/write performance without considering the effect of directory layout. It has been found that the performance of the file system operations is heavily dependent on the hardware architecture and corresponding parameters like bus speed, bus width, memory size, protocol, etc. However if these parameters are held constant then the directory depth, where a file is located, becomes a major parameter influencing the performance of a file system. With the explosive growth of data to be stored, we need better metadata management techniques to improve the accessibility of actual contents.

The present day popular file systems create and maintain directory files in the same manner as the data files are kept. Such approach allows the directory files to be placed randomly over the disk and receive non-contiguous space, resulting in increased overhead while resolving a path. However, existing file systems are flexible enough to accommodate different customized layouts of file storage by different types of users. With the help of optimization techniques used by the operating system, the implementation of storage structures is kept transparent to the user. In terms of workstations, file access is usually a step-by-step procedure which is very much in the context of usage, allowing optimizations to take place. On the other hand in servers, request for a file from the storage medium is serviced which is generally out-of-context. Such a request needs to traverse the whole path in stages where no optimization technique like caching is effective, resulting in slow response.

The study presented in this paper analyzes the effect of directory depth over the file system efficiency. We analyzed the file system performance for various directory depths while pertaining to the file system design. We developed a benchmark application that emphasize on the parameters required for the file accessing behaviors of the file systems. In [4], Tanenbaum et al. argue that most of the prominent file system architectures are hierarchical in nature. In our studies, we analyzed the overhead involved in data read operations due to hierarchical directory layout in most common file systems.

A. Related work

The File layout and file system performance is studied in [3]. They found that there is a sifnificant performance degradation due to fragmented files on the storage media. The file size distribution on UNIX is studied by A.S. Tanenbaum and others in [4]. The file system space utilization is presented by [5], [6]. The File system usage in Windows NT 4.0 is presented in [7]. The authors present their results about parameters like file life time, data distribution, file access patterns, file opening and closing characteristics etc. Furthermore, there are many different benchmarking applications for file systems as available in [8], [9], [10], [11]. The performance impact

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

of stripe size on the network attached storage systems is presented in [12]. Metadata management for large scale file systems is presented in [13]. The metadata indexing and search in petascale data storage systems is given in [14]. In this manuscript, we contribute by presenting our results for the panelty due to hierarchical nature of file systems found through our benchmark application. This studies shares the main objective of improving the file system performance by studying the existing file systems with other related work given above.

B. Overview

In this paper we present our findings about the effect of directory depth over performance of FAT-32, NTFS, UFS and EXT-2 file systems. In section 2, we give a short description of our Benchmark application, section 3 presents the environment and settings during data collection and in section 4, we describe our results with discussions. Section 5 presents the conclusion and future illusions.

II. Benchmark description and data collection

The following are the major components of our benchmarking process.

- Workload Generator
- Supervisor
- Client Application

A. Workload Generator

The workload is created on a logical partition of the storage medium on the server. The Workload Generator program creates files and directories on the storage media, which are used for collection of results. Since, in describing our results, different levels in the directory hierarchy will be periodically referred, therefore to avoid any confusion we will refer the root level as level-0, a directory on the root as level-1, subdirectory within a directory on the root as level-2, and so on; as shown in figure 1. In our experimental set up, the data was generated to 15 levels down the hierarchy. The workload taken into consideration consisted of 32,000 files (32,000 is upper limit of number of files in FAT32) and a similar number of directories at the root level of the logical partition of storage media i.e. level-0. For levels down the hierarchy, each sub-directory contains one file and one directory, which further has a file and a directory and so on. Thus, each level, including the root level contains 32,000 files and a similar number of directories. This workload can be increased to any number of levels, depending upon the partition size of the hard disk and cluster size being used by the operating system.

B. Supervisor

The supervisor is used for supervision of the client components. It issues a few supervisory commands to clients as described below.

- Clients need some initializing parameters which are sent by the supervisor. In case of multiple clients, supervisor is capable of sending the same parameters with varying values, dynamically.
- The set of parameters include the number of requests to be sent from each client, the directory level for accessing files and a unique seed to generate a different random path for the files at the same directory level.
- When a client sends a registration request, the supervisor adds it in the List of Clients and sends back the above parameters on the basis of which client will access random files located on the server.
- Supervisor multicasts a go command when the start button is pressed. Thus, all the clients start their corresponding data read operation from files located at the Server, simultaneously.
- After completion of the requests, all clients will send their results to the supervisor. Once all results are collected, the supervisor makes a log file when the Save Results button is pressed.

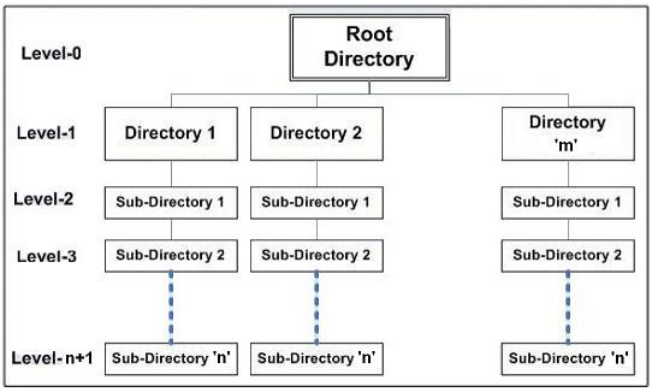


Fig. 1. Sample Data Layout Created via Workload Generator

C. Client Application

The clients send requests to the file server to read from random paths and pursue the following steps:

- When a client is initialized, it notifies the Supervisor for its existence.
- In response, the supervisor sends back set of parameters, on the basis of which client application sets the values of different parameters required for benchmarking.
- Client waits until it receives a go command from the Supervisor application
- It starts sending a specified number of requests to the file server on the go command.
- The Client reads the first byte of each request from a unique file at specified depth.
- After completing the requests, client sends results to the supervisor.

The communication between the Supervisor and the Clients contains a sequence of parameters. This sequence is initiated by the clients request to register itself at the Supervisor. In response, the Supervisor generates a random number based on the client-ID and the time at which request is received and sends it to the client with other essential parameters. Thus,

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

in response to its registration request, the client gets a wait signal and a random number. The clients use the random number to generate a list of files from a configuration file which contains paths to all the files on the server with respect to the particular directory level. On completion of the read requests, each client sends the elapsed time for the request to the supervisor. In reply, the supervisor again sends a random number as a seed for the next requests, on the basis of which the client regenerates a new list of files to access and waits for the next go command.

III. Experimental environment and configurations

The studies presented in [5], [6], show the usage patterns of data in different working environments. For our observations, we generated data according to the findings in these papers. The data was quite fragmented on the hard drive to minimize the automatic optimizations. Additionally, traditional measures of performance optimizations for the working of a file system, like caching, were tried to be reduced by disabling them to acquire the actual overheads at different directory depths.

We read only first byte of each file to make the file access time constant for each directory level. Furthermore, it kept seek time for a file to be constant. Therefore, as a result the variable time showed the latency due to the path traversal in the file systems. The cluster size for the file systems was kept to its default level.

The systems, we used for our experiments had the same hardware configurations for Windows and Linux based volumes and were connected through a 100-bit Ethernet LAN. The hard disk drives were of 40GB capacity, out of which a 20GB partition was used for data generated by the application as in section II-A. Ten clients of the same hardware configuration and operating system were set up. All clients initially registered to a single supervisor as in section II-B on the network. Each client sent 1000 requests to the server by randomly selecting from the configuration file, i.e. a total of 10,000 requests were sent at each directory level on the server. Zero depth of directory shows that the file being requested by the client was physically present on the root level of logical partition. The File server was bombarded with a chain of read requests via certain client-nodes over the network. This chain of requests was repeated for different directory depths. We performed our experiments separately for all the file systems i.e. FAT32, NTFS and EXT. During our experiments, we used a dedicated network, thus there was not any other network traffic. All the file servers had data generated from the same configuration file which means that all of them had same number of files and directories at same directory level with respect to their root level directory.

In our client settings, we had 10 clients with similar hardware configurations. Prior to start the requests for data from the file server, clients (as described in II-C) got registered with a monitor for reporting errors or final results otherwise. We copied all the file paths to the clients in a file.

TABLE I
Penalty in FAT32 File System Levels

Directory Level	Avg. Time (seconds)	Penalty Ratio w.r.t. Level-0
0	121	1
1	350	2.89
2	405	3.35
3	442	3.65
4	464	3.83
5	486	4.02
6	505	4.17
7	524	4.33
8	557	4.6
9	598	4.94
10	632	5.22
11	658	5.44
12	691	5.71
13	709	5.86
14	736	6.08
15	757	6.26

A. Windows based Volumes

We completed our experiments for the FAT32 and NTFS file systems in the same environment. The cluster block size for both the file systems in use was kept to their default values (which are generally used), i.e. 4 KB for NTFS and 16 KB for FAT 32 on a logical partition of 20 GB. The size of the data generated on the disk drive formatted with FAT-32 file system is 19.6GB and the logical drive with NTFS format is populated with 7GB of data. To avoid any performance increasing mechanism, caching client requests for shared data parameter was disabled.

B. Linux and Solaris based volmues

We used a dedicated sun system as file server with UFS. Similarly, we had a system with 3GHz processor, 1GB memory and 40GB disk drive as a file server with EXT-2 file system. Clients accessed the file servers through ”SAMBA” utility for file sharing.

It should be noted that for our observations for UFS, we used Sun Solaris systems which had different hardware configurations but generated data was of similiar nature as of others. Therefore, we donot compare the results.

IV. Results and discussion

We collected results for all FAT-32, NTFS and EXT-2 file systems in almost similar environment. The environment for UFS was different as it uses Sun Solaris systems but it is not a matter as we did not intend the comparision of the performance between different file systems. Our objective was to explore the performance degradation along the directory depth.

Preliminary, experiments were performed before choosing the final values of the three main variable parameters: number of clients, number of requests, and directory depth. An increasing number of clients increase the network traffic, resulting in more collisions. The time to retransmit the request affects the measurements. Same effect is observed in terms of increase in the number of requests. We minimized the chances of caching on the server by using a moderate number of requests. We set directory depth at fifteen levels, starting from level 0 which

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

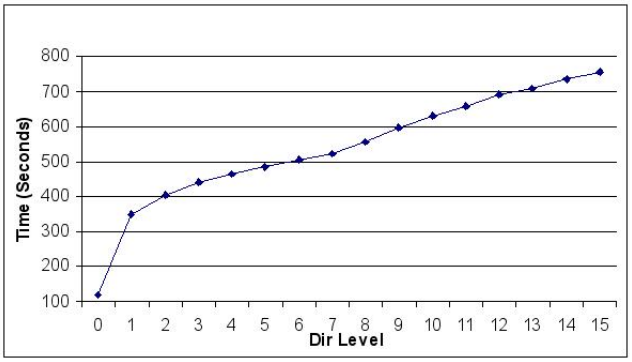


Fig. 2. FAT32 10 Clients and 1000 Requests/Client

is the root directory as observed in [5], [6]. Figure 2 shows the results for each directory level, plotted against time taken for a thousand requests from each client for the FAT32 file system. An average of five readings for the time taken at each directory level from 10 clients is plotted against each directory level. There is an abrupt increase in file access time when going from root level to the first level of depth. After this, there is a linear increase found in the observations on FAT-32 file system.

Table I shows that the penalty for each directory other than the root directory, i.e. level 0, increases significantly. For an operation on a file located in the first subdirectory or level 1, it takes about three times as long and for the level-10 it is more than five times as long as the time taken at the root directory. The difference of penalty ratios of two adjacent levels is somewhat consistent; except that of level 0 and 1. Study of the FAT-32 file system architecture shows that it implements a linked based allocation scheme and every path traversal starts from the root directory which is treated as a reference point [14]. For this purpose the root directory is cached at system startup. Therefore, for any file located on the volume the file system will start from the root directory and traverse the path step by step before the file is located. The same experiment with similar parameters was carried out

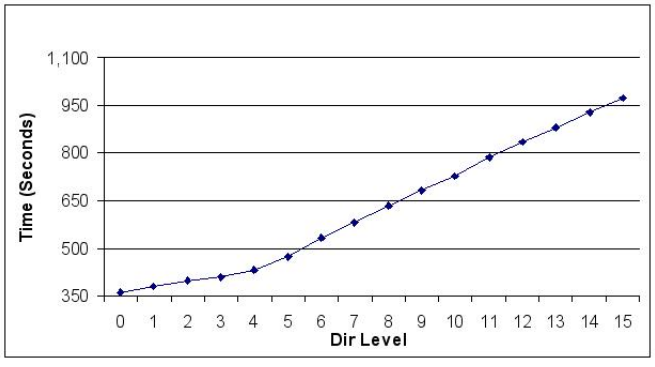


Fig. 3. NTFS 10 Clients and 1000 Requests/Client

on the NTFS-based volume. Corresponding results of NTFS based experiments are plotted in figure 3. The curve in the graph shows the effect of directory depth on the file system performance in NTFS. It is also evident that the increase in time up to fourth level in the NTFS file system is less than the increase at higher directory levels. Our results show that from level 5 and onwards the increase in penalty with respect to the previous level is more than 10%. However, by level 10 file access penalty is doubled than that of the root directory and, for further directory depths, it continues to increase constantly. This less significant increase at the initial levels is due to the structure of NTFS file systems. The structure of NTFS shows that it uses master file tables (MFT) for managing data. The study of the structure of the NTFS file system conducted in [15] reveals that up to some directory depth data is stored directly in MFT which decreases the access time for initial levels of directory depth.

Table II shows that for the first four levels the files were stored in MFT, but for higher levels a hierarchical directory structure was chosen for storage of files. It is also evident from the table that the difference of penalty ratios for any two adjacent levels is less till level 4, but increases there after. In

TABLE II
PENALTY IN NTFS FILE SYSTEM LEVELS

Directory Level	Avg. Time (seconds)	Penalty Ratio w.r.t. Level-0
0	362	1
1	378	1.04
2	397	1.1
3	411	1.14
4	433	1.2
5	475	1.31
6	532	1.47
7	583	1.61
8	634	1.75
9	683	1.89
10	726	2.01
11	786	2.17
12	836	2.31
13	880	2.43
14	928	2.56
15	972	2.69

Figures 4, 5, we show the results observed for EXT-2 and UFS file systems. It is seen that graphs are more linear in case of EXT-2 and UFS file systems. We see that there is a steady increase in penalty along the increase in directory depth. The performance degradation is obvious. Tables III and IV display

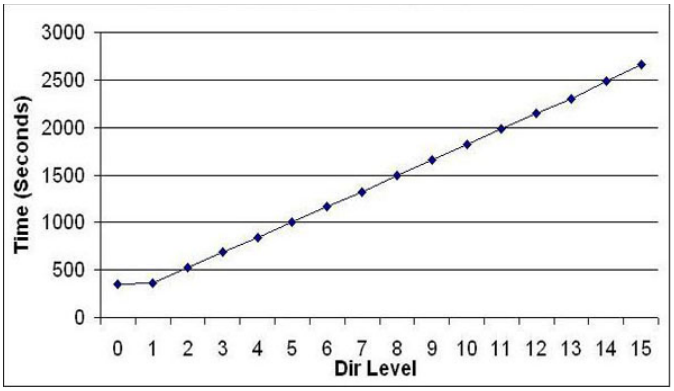


Fig. 4. EXT-10 Clients- 1000 Requests/Client

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

TABLE III
PENALTY IN EXT-2 FILE SYSTEM LEVELS

Directory Level	Avg. Time (seconds)	Penalty Ratio w.r.t. Level-0
0	345	1.00
1	360	1.04
2	522	1.51
3	685	1.98
4	837	2.42
5	1000	2.90
6	1168	3.38
7	1320	3.82
8	1492	4.32
9	1654	4.79
10	1821	5.27
11	1981	5.74
12	2142	6.21
13	2299	6.66
14	2486	7.20
15	2657	7.70

TABLE IV
PENALTY IN UFS FILE SYSTEM LEVELS

Directory Level	Avg. Time (seconds)	Penalty Ratio w.r.t. Level-0
0	117	1.00
1	148	1.26
2	185	1.58
3	220	1.87
4	255	2.18
5	290	2.48
6	325	2.77
7	364	3.11
8	393	3.36
9	434	3.70
10	468	3.99
11	499	4.26
12	532	4.54
13	573	4.89
14	609	5.20
15	644	5.50

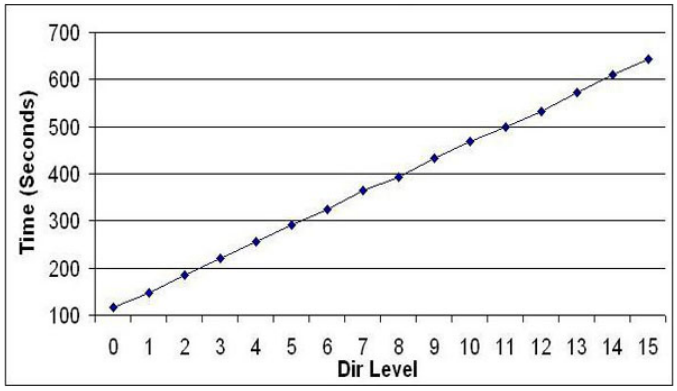


Fig. 5. UFS-10 Clients- 1000 Requests/Client

the raw results for further calculations. We can calculate the penalty factor F from these values.

The purpose of this study was to explore the time taken in accessing files located at different levels of the directory hierarchy. We did not investigate about the loopholes present in the existing file systems. Furthermore, we did compare different file systems in terms of using better optimizing techniques. The objective of the conducted experiment was purely to analyze relative performance of different file systems in terms of speed of file accessing at different levels. For this purpose, known optimizing techniques like caching were explicitly disabled during the data collection process. Thus, the only factor investigated was the operating expense of directory traversal at different levels.

V. CONCLUSION AND FUTURE WORK

This paper gives a brief overview of the effect of directory depth on file system efficiency. Our results showed that there is a significant increase in access time with increasing directory depth. We noted that the performance of the file system was mainly affected by the directory depth where many clients access files from different directories, which is a fundamental requirement for a file server. Benchmark results showed that a linear but significant increase was found in terms of time as we accessed a file at deeper directory level. The results shown in table II indicate that the difference in percent increase of time is approximately 70 sec, amongst levels 5, 10 and 15 on NTFS volumes. Similarly, results in the FAT-32 file system, shown in table-2, indicate that an average difference in percent increase of file access time is approximately 112 sec between levels 5, 10 and 15. Similiarly, a considerable performance degration is observed in EXT-2 and UFS file systems as show in tables III, IV.

In this manuscript, we have presented our observations for different file systems. The larger objective of our work is to develop a more efficient and flexible design of storage media. The idea is to create a data server based on factual data patterns. The benchmarking results presented here reveal that there is a strong need to consider new techniques for data management on storage media. We propose a new file system where metadata should be completely separated from the data on the disk drives. This will decrease this perfomance overhead due to dispersed directories over the disk drives.

VI. ACKNOWLEDGMENTS

Authors would like to acknowledge support provided by the Higher Education Commission Islamabad, Pakistan through the Indigenous Ph.D. Fellowship Program for conducting this research.

REFERENCES

[1] Oracle, “Linux file system performance comparison for oltp with ext2, ext3, raw, and ocfs on direct-attached disks using oracle 9i release 2,” January 2004.

[2] F. Ahsan, M. I. Lali, I. Ahmad, A. F. M. Ishaq, and S. Mohsin, “Exploring the effect of directory depth on file access for FAT and NTFS file systems,” in *ISTASC’08: Proceedings of the 8th conference on Systems theory and scientific computation*. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS), 2008, pp. 130–135.

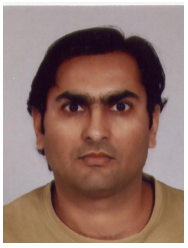
[3] K. A. Smith and M. Selzter, “File layout and file system performance,” Harvard Computer Science, USA, Technical Report TR-35-94, 1994.

[4] A. S. Tanenbaum, J. N. Herder, and H. Bos, “File size distribution on unix systems: then and now,” *Operating Systems Review*, vol. 40, no. 1, pp. 100–104, 2006.

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

[5] M. I. Ullah, F. Ahsan, and A. F. M. Ishaq, "Study of File System Space Utilization Patterns in MS-Windows Volumes," *Proceedings of International Conference on ICT in Education and Development*, pp. 157–164, December 16-18th, 2004.
[6] M. I. Ullah, F. Ahsan, I. Ahmad, and A. F. M. Ishaq, "Analysis of File System Space Utilization Patterns in UNIX Based Volumes," *Proceedings of The IEEE International Conference on Emerging Technologies, (ICET 2005)*, Septemebr 17-18, 2005.
[7] Vogels and Werner, "File system usage in windows nt 4.0," in *SOSP '99: Proceedings of the seventeenth ACM symposium on Operating systems principles*. New York, NY, USA: ACM, 1999, pp. 93–109.
[8] T. Bray, "Bonnie file system benchmark," [Online], Available: http://www.textuality .com / bonnie/. (Last visited 01/10/2009), November 1990.
[9] F. John, "Whos best? how good are they? how do we get that good?" [Online], Available: http://management.about.com/cs/benchmarking/a/-Benchmarking.htm (Last visited 11/10/2009), November.
[10] J. B. A. Park, "Iostone: A synthetic file system benchmark," pp. 45–52, June 1990.
[11] Iozone, "Iozone filesystem benchmark," [Online], Available: http://www.iozone.org/, July, 2005 bonnie/. (Last visited 07/10/2009), July 2005.
[12] Y. Deng and F. Wang, "Exploring the performance impact of stripe size on network attached storage systems," *Journal of Systems Architecture*, vol. 54, no. 8, pp. 787–796, 2008.
[13] S. Weil, S. A. Brandt, E. L. Miller, and K. Pollack, "Intelligent metadata management for a petabyte-scale file system," May 2004.
[14] A. Leung, M. Shao, T. Bisson, S. Pasupathy, and E. L. Miller, "High-performance metadata indexing and search in petascale data storage systems," Jul. 2008.
[15] D. Mikhailov, "FAT and NTFS performance," [Online], Available:http://www.digit-life.com/articles/ntfs/index3.html (Last visited 15/09/2009), June 2007.

Author's Profile



Mr. M.I. Lali received his master in software engineering degree from COMSATS Institute of Information Technology (CIIT), Islamabad, Pakistan in 2002. Later, after spending a few years in industry, he joined CIIT for PhD program. Currently, he is pursuing his studies for PhD in the area of Formalism in File Systems and Software Design. He has worked at University fo Groningen, Netherlands for some time for his research.

Mr. F. Ahsan received his Bachelors in Computer Science degree from FAST, National University, Karachi, and masters degree from SZABIST, Islamabad, Pakistan. Currently, he is pursuing his PhD studies at COMSATS Institute of Information Technology, Islamabad. His research is focused on distributed systems and computer networks, supported by HEC.

Dr. Ishaq earned a doctorate in Physics from McMaster University, Hamilton, Canada in 1972. He switched over to Computer Systems in the late seventies. He has 43 years of professional experience in university teaching, research, training, academic administration, technical consulting and management. His last full time employment was with CIIT, Islamabad, as a Professor and Dean. He moved to Dubai, UAE, in 2007, where he works as a Consultant and is associated with SZABIST in Dubai International Academic City.

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

Using Statistical Moment Invariants and Entropy in Image Retrieval

Ismail I. Amr*, Mohamed Amin†, Passent El-Kafrawy†, and Amr M. Sauber†

*College of Computers and Informatics
Misr International University, Cairo, Egypt
†Faculty of science Department of Math and Computer Science
Menoufia University, Shebin-ElKom, Egypt

Abstract

Although content-based image retrieval (CBIR) is not a new subject, it keeps attracting more and more attention, as the amount of images grow tremendously due to internet, inexpensive hardware and automation of image acquisition. One of the applications of CBIR is fetching images from a database. This paper presents a new method for automatic image retrieval using moment invariants and image entropy, our technique could be used to find semi or perfect matches based on query-by-example manner, experimental results demonstrate that the purposed technique is scalable and efficient.

Keywords

Moment invariants, content-based image retrieval, image entropy.

1 Introduction

In many areas of commerce, government, academia, and hospitals, large collections of digital images are being created. Many of these collections are the product of digitizing existing collections of analogue photographs, diagrams, drawings, paintings, and prints. Usually, however, technologies related to archiving, retrieving, and editing images/video based on their content are still in their infancy, the only way of searching these collections was by keyword indexing, or simply by browsing. Digital image databases however, open the way to content-based searching. ”Content-based” means that the search will analyze the actual contents of the image. The term content’ in this context might refer to color, shape, texture, or any other information that can be derived from the image itself. Without the ability to examine image content, retrieval must rely on metadata such as captions or keywords, which may be laborious or expensive to produce.

What is desired is a similarity matching, independent of translation, rotation, and scale, between a given template (example) and images in the database. Consider the situation where a user wishes to retrieve all images containing cars, people, etc. in a large visual library. Being able to form queries in terms of sketches, structural descriptions, color, or texture, known as query-by-example (QBE), offers more flexibility over simple alphanumeric descriptions. This paper is organized as follows: section 2 provides the necessary background for CBIR. Section 3 defines the image segmentation technique using the Moments and Entropy. Section 4 explains the pro- posed model with the experiments conducted. Finally, section 5 concludes the paper and lists some further work.

1.1 Background

Among the approaches used in developing early image database management systems (IDMS) are textual encoding [1], logical records [2], and relational databases [3]. The descriptions, employed to convey the content of the image, were mostly alphanumeric. Furthermore, these were obtained manually or by utilizing simple image processing operations designed for the application at hand. Later generations of IDMS have been designed in an object-oriented environment [4], where image interpretation routines form the backbone of the system. However, queries still remain limited to a set

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

of predetermined features that can be handled by the system. The reader is referred to [5] for a survey of IDMS.

Most recent systems reported in the literature for searching, organizing, and retrieving images based on their content include IBMs Query-by-Image-Content (QBIC) [6], MITs photo-book [7], the Trademark and Art Museum applications from ETL [8], Xenomania from the University of Michigan [9], and Multimedia/VOD test bed applications from the Columbia University [10]. IBMs QBIC is a system that translates visual information into numerical descriptors, which are then stored in a database. It can index and retrieve images based on average color, histogram color, texture, shape, and sketches. MITs photobook describes three content-based tools, utilizing principal component analysis, finite element modes, and the Wold transform to match appearances, shapes, and textures, respectively, from a database to a prototype at run time. Xenomania is a system for face image retrieval, which is based on QBE. Its embedded routines allow for segmentation and evaluation of objects based on domain knowledge, yielding feature values that can be utilized for similarity measures and image retrieval. The database management system of the Columbia University proposes integrated feature maps based on texture, color, and shape information for image indexing and query in transform domain.

Similarity-based searching in medical image databases has been addressed in [11]. A variety of shape representation and matching techniques are currently available, which are invariant to size, position, and/or orientation. They may be grouped as: (1) methods based on local features such as points, angles, line segments, curvature, etc. [12]; (2) template matching methods [13]; (3) transform coefficient based methods, including Fourier descriptors [14] or generalized Hough transform [15]; (4) methods using 3 modal and finite element analysis [16]; (5) methods based on geometric features, such as local and differential invariants [17]; and (6) methods using B-Splines or snakes for contour representation [18]. Comprehensive surveys of these methods can be found in [19].

III. BACKGROUND

A. Image Segmentation

The shape representation method described here assumes that the object has been fully segmented from the original image, such that all pixels representing the objects shape have been identified as distinct from those pertaining to the rest of the image. In this paper, a local diffusive segmentation method [20] is used. There exist a wide variety of ways to achieve segmentation; however, it is not the subject of this paper. All contiguous pixels, which share a given point-based characteristic of the object or are surrounded by those that do, are considered as object pixels and those outside the included region, are considered as background. The result is a group of contiguous pixels, which collectively represent the object. The boundary pixels of the object are then extracted from the segmented object pixels by a simple iterative trace, around the outside of the object that continues until the starting point is reached. This trace produces a second group of pixels collectively representing the objects exterior contour [20].

B. Entropy

Entropy is a scalar value representing a statistical measure of randomness that can be used to characterize the texture of the input image. Entropy is defined as

$$S = -\sum_{i=1}^{\Omega} P_i \log_2(P_i)$$

The value of entropy is also an invariant that is neither affected by rotation nor scaling.

C. Moments

Region moment representations interpret a normalized gray level image function as a probability density of a 2D random variable. Properties of this random variable can be described using statistical characteristics - moments. A moment of order (p+q) is dependent on scaling, translation, rotation, and even on gray level transformations and is given by

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q g(x,y)\,dx\,dy$$

The central moment

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-x_c)^p (y-y_c)^q g(x,y)\,dx\,dy$$

Let $f(x,y) = (x-x_c)^p (y-y_c)^q g(x,y)$

We use composite trapezoidal rule for evaluation the double integral

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

$$I_T = \int_{-\infty}^{M}\int_{-\infty}^{N} f(x,y)dxdy$$

$$I_T = \frac{hk}{4}\begin{bmatrix} \{f_{0,0} + f_{0,M} + 2(f_{0,1} + f_{0,2} + ... + f_{0,M-1})\} \\ +2\sum_{i=1}^{N-1}\{f_{i,0} + f_{i,M} + 2(f_{i,1} + f_{i,2} + ... + f_{i,M-1})\} \\ \{f_{N,0} + f_{N,M} + 2(f_{N,1} + f_{N,2} + ... + f_{N,M-1})\} \end{bmatrix}$$

Where M and N image size, x_c , y_c are the co-ordinate of the region's centriod , $h = k = 1$ and

$$y_c = \frac{m_{01}}{m_{00}},\ x_c = \frac{m_{10}}{m_{00}}$$

The normalized un-scaled central moments

$$\vartheta_{pq} = \frac{\mu_{pq}}{(\mu_{00})^{\lambda}}$$

Where $\lambda = \frac{P+q}{2} + 1$

A less general form of invariance is given by seven rotation, translation, and scale invariant moment characteristics [31].

$$\phi_1 = \vartheta_{20} + \vartheta_{02}$$

$$\phi_2 = (\vartheta_{20} - \vartheta_{02}) + 4\vartheta_{11}^2$$

$$\phi_3 = (\vartheta_{20} - 3\vartheta_{12})^2 + (3\vartheta_{21} - \vartheta_{03})^2$$

$$\phi_4 = (\vartheta_{30} - \vartheta_{12})^2 + (\vartheta_{21} - \vartheta_{03})^2$$

$$\phi_5 = \begin{bmatrix} (\vartheta_{30} - 3\vartheta_{12})(\vartheta_{30} + \vartheta_{12})\left[(\vartheta_{30} + \vartheta_{12})^2 - 3(\vartheta_{21} + \vartheta_{03})^2\right] \\ +(3\vartheta_{21} - \vartheta_{03})(\vartheta_{21} + \vartheta_{03})\left[3(\vartheta_{30} - \vartheta_{21})^2 - (\vartheta_{21} + \vartheta_{03})^2\right] \end{bmatrix}$$

$$\phi_6 = \begin{bmatrix} (\vartheta_{20} - \vartheta_{02})\left[(\vartheta_{30} + \vartheta_{12})^2 - (\vartheta_{21} + \vartheta_{03})^2\right] \\ +4\vartheta_{11}(\vartheta_{30} + \vartheta_{12})(\vartheta_{21} + \vartheta_{03}) \end{bmatrix}$$

$$\phi_7 = \begin{bmatrix} (3\vartheta_{21} - \vartheta_{03})(\vartheta_{30} + \vartheta_{12})\left[(\vartheta_{30} + \vartheta_{12})^2 - 3(\vartheta_{21} + \vartheta_{03})^2\right] \\ -(\vartheta_{30} - 3\vartheta_{12})(\vartheta_{21} + \vartheta_{03})\left[3(\vartheta_{30} + \vartheta_{21})^2 - (\vartheta_{21} + \vartheta_{03})^2\right] \end{bmatrix}$$

2 The Proposed Technique

Our proposed method consists of two parts: region selection and shape matching. In the first part, the image is partitioned into disjoint, connected regions. The second part, the shape of each potential object is tested to determine whether it matches one from a set of given template. To this effect, we use image Moments and Entropy. Although many techniques [21]–[23] use moment invariants to overcome rotation and scaling, we extend this technique by using both moments and entropy to support two level filtering which is more appropriate in an image database case. In this paper we focus on part two and part one may be implemented as shown in section III-A.

A. Implementation

Images, as figure 1, are stored in the database after its subdivided into distinct sub-images as stated previously. For each sub-image stored we compute the entropy and moments, that is indexed accordingly.

Each time a template image is requested to find its match in the database, the following steps are implemented. First we compute the entropy and moments for the image that is being searched for. Second, the computed entropy is used to filter the images in the database. At last, the computed moments are used to determine and fetch the most matching images in the filtered subset.

B. Experimental Results

The experiments run on an image database that contains one hundred images. Figure 1 shows all the images, where each of these sub-images are stored independently in the database. We picked some images in the database scaled, rotated, or scaled and rotated and used as testing templates. Figure 2 and 3 show multiple experimental results that demonstrate the performance of our technique, each experiment contains a template (example) and a result set containing the matches. All the results are promising except when the template is scaled larger than double the size of the original image (i.e. image in the database).

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

Figure 1 Images in database

Experiments

Template	Result
	
	

Conclusions

This paper presents a method for automatic CBIR based on query-by-example by region-based shape matching. The proposed method consists of two parts: region selection and shape matching. Each image is segmented to a set of sub images by using a local diffusive segmentation method. Consequently, the image entropy is computed and used to narrow the search space. Later, the moment invariants of the image are matched, which are independent to translation, scale, rotation and contrast, to every sub image and to the template. Retrieval is performed by returning those sub images whose invariant moments are most similar to the ones of a given query image.

In the future we would like to find more image characteristics that can be used to query images in the database. Our focus now is to explore other characteristics that can be used for highly scaled images. Moreover, we will research colored images and develop new techniques to handle color.

ACKNOWLEDGMENT

The author would like to thank the valuable comments and encouragement received from colleagues in Menoufia University.

REFERENCES

[1] S. K. Chang, “Pictorial information systems,” IEEE Computer, vol. 14, no. 11, 1981.

[2] W. I. Grosky, “Toward a data model for integrated pictorial databases,” Computer Vision Graphics Image Process, vol. 25, no. 3, pp. 371–382, 1984.

[3] N. S. Chang and K. S. Fu, “Picture query languages for pictorial database systems,” IEEE Computer, vol. 14, no. 11, pp. 23–33, 1981.

[4] A. C. A. Pizano, A. Klinger, “Specification of spatial integrity constraints in pictorial databases,” IEEE Computer, vol. 22, no. 12, pp. 59–70, 1989.

[5] W. I. Grosky and R. Mehrotra, “Image database management,” Advances in Computers, vol. 35, no. 5, pp. 237–291, 1992.

[6] W. N. J. A. Q. H. B. D. M. G. J. H. D. L. D. P. D. S. P. Y. M. Flickner, H. Sawhney, “Query by image and video content: The qbic system,” Computer, vol. 28, no. 9, pp. 23–32, 1995.

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

[7] S. S. A. Pentland, R. W. Picard, “Photobook: Content-based manipulation of image databases,” in Proc. SPIE Storage and Retrieval for Image and Video Databases II, W. N. R. C. Jain, Ed., vol. 2, no. 185, SPIE. Bellingham, Wash., 1994, pp. 34–47.
[8] H. S. T. Kato, T. Kurita, “Intelligent visual interactionwith image database systems-toward multimedia personal interface,” Journal of Information Processing, vol. 14, no. 2, pp. 134–143, 1991.
[9] R. J. J. R. Bach, S. Paul, “A visual information management system for the interactive retrieval of faces,” IEEE Trans. Knowledge Data Engineering, vol. 5, pp. 619–628, 1993.
[10] J. R. Smith and S. Chang, “Visualseek: A fully automated content-based image query system,” in ACM Multimedia Conference, Boston, MA, 1996.
[11] C. F. E. G. M. Petrakis, “Similarity searching in medical image databases,” University of Maryland, Tech. Rep. UMIACS-TR-94-134, 1994.
[12] F. Stein and G. Medioni, “Structural indexing: Efficient 2-d object recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 14, pp. 1198–1204, 1992.
[13] W. J. R. D. P. Huttenlocher, G. A. Klanderman, “Comparing images sing the hausdorff distance,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 15, pp. 850–863, 1993.
[14] P. Kumar and E. Foufoula-Georgiou, “Fourier domain shape analysis methods: A brief review and an illustrative application to rainfall area evolution,” Water Resources Res., vol. 26, pp. 2219–2227, 1990.
[15] S. C. Jeng and W. H. Tsai, “Scale and orientation invariant generalized hough transform—a new approach,” Pattern Recognit., vol. 24, no. 11, pp. 1037–1051, 1991.
[16] S. Sclaroff and A. P. Pentland, “Modal matching for correspondence and recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 17, pp. 545–561, 1995.
[17] E. Rivlin and I. Weiss, “Local invariants for recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 17, pp. 226–238, 1995.
[18] Z. Y. F. S. Cohen, Z. Huang, “Invariant matching and identification of curves using b-splines curve representation,” IEEE Trans.Image Process., vol. 4, pp. 1–10, 1995.
[19] W. E. L. Grimson, Object recognition by computer: The role of geometric constraints. Cambridge, MA: MIT Press, 1990.
[20] T. Bernier, “An adaptable recognition system for biological and other irregular objects,” Ph.D. dissertation, McGill University, 2001.
[21] J. Flusser, “On the independence of rotation moment invariants,” Pattern Recognition, vol. 33, no. 9, pp. 1405–1410, 2000.
[22] S. M. S. Rodtook, “Numerical experiments on the accuracy of rotation moments invariants,” Image and Vision Computing, vol. 23, no. 6, pp. 577–586, June 2005.
[23] H. L. Dong Xu, “Geometric moment invariants,” Pattern Recognition, vol. 41, no. 1, pp. 240–249, January 2008.

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

Genetic Algorithm Based Optimization of Clustering in Ad-Hoc Networks

*[1]Bhaskar Nandi #[2]Subhabrata Barman #[3]Soumen Paul

#[2,#3]*Haldia Institute of Technology, W.B, India.*
#2

*1

***Abstract**:*
In this paper, we have to concentrate on implementation of Weighted Clustering Algorithm with the help of Genetic Algorithm (GA).Here we have developed new algorithm for the implementation of GA-based approach with the help of Weighted Clustering Algorithm (WCA) [4]. Cluster-Head chosen is a important thing for clustering in ad-hoc networks. So, we have shown the optimization technique for the minimization of Cluster-Heads(CH) based on some parameter such as degree-difference , Battery power (P_v), degree of mobility, and sum of the distances of a node in ad-hoc networks. Cluster-Heads selection of ad-hoc networks is an important thing for clustering. Here, we have discussed the performance comparison between deterministic approach and GA-based approach. In this performance comparison, we have seen that GA does not always give the good result compare to deterministic WCA algorithm. Here we have seen connectivity (connectivity can be measured by the probability that a node is reachable to any other node.) is better than the deterministic WCA algorithm [4].

***Keywords*- Adhoc Networks, GA, Cluster Head (CH), WCA.**

I. INTRODUCTION

A wireless ad-hoc network consists of nodes that move freely and communicate with each other using wireless links. Ad-hoc networks do not use specialized routers for path discovery and traffic routing. One way to support efficient communication between nodes is to develop wireless backbone architecture; this means that certain nodes must be selected to form the backbone. Over time, the backbone must change to reflect the changes in the network topology as nodes move around. The algorithm that selects the members of the backbone should naturally be fast, but also should require as little communication between nodes as possible, since mobile nodes are often powered by batteries. One way to solve this problem is to group the nodes into clusters, where one node in each cluster functions as cluster head, responsible for routing. A clusterhead does the resource allocation to all the nodes belonging to its cluster. Due to the dynamic nature of the mobile nodes, their association and dissociation to and from clusters perturb the stability of the network and thus reconfiguration of cluster heads is unavoidable. Thus, it is desirable to have a minimum number of clusterheads that can serve the network nodes scattered evenly in the area. An optimal selection of the clusterheads is an NP-hard problem . Therefore, various heuristics have been designed for this problem . we apply genetic algorithms (GA) as an optimization technique to improve the performance of clusterhead election procedure. In particular, GAs are defined as search algorithms that use the mechanics of natural selection and genetics such as reproduction, gene crossover, mutation as their problem-solving method. The goal is to he able to find out a better solution in the form of new generations that have received advantages and survival-enhancing traits from the previous Generations. We have to target artificial-life simulation is created where survival of the fittest logic is applied for the string structures that are the living organism equivalent in real world. Even though the representation is structured, there is a randomization in data exchange to simulate the evaluation of real life forms. As each generation brings up a new set of strings by different combination of bits of pieces of the previous generation, the results are not guaranteed to come up with a generation that has a better fitness value hut by performing different genetic operations, the probability of achieving the desired results is increased.

II. CLUSTERING IN ADHOC NETWORKS

The weight-based distributed clustering algorithm that takes into consideration that the number of nodes that a cluster head can handle the ideal degree, transmission power, mobility and battery power of a mobile node. We try to keep the number of nodes in a cluster around a pre-defined threshold to facilitate the optimal operation of the medium access control (MAC) protocol. Our cluster head election procedure is n periodic as in earlier research, but adapts based on the dynamism of threshold value of nodes. This on-demand execution of WCA aims to maintain the stability of the network, thus lowering the computation and communication cost associated with it.

A cluster head may not be able handle a large number of nodes due to resource limitations even if these nodes are its neighbors and lie well within its transmission range. Thus, the load handling capacity of the cluster head puts an upper bound on the node-degree. In other words, simply covering the area with the minimum number of cluster heads will put more burden on the cluster heads. At the same time, more cluster heads will lead to a computationally expensive system. This may result in good throughput, but the data packets have to go through multiple hops resulting in high latency. In summary, choosing an optimal number of cluster heads which will yield high throughput but incur as low latency as possible, is still an important problem. As the search for better heuristics for this problem continues, we propose the use of a combined weight metric, that takes into account several system parameters like

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

the ideal node-degree, transmission power, mobility and the battery power of the nodes. We could have a fully distributed system where all the nodes share the same responsibility and act as cluster heads. However, more cluster heads result in extra number of hops for a packet when it gets routed from the source to the destination, since the packet has to go via larger number of cluster heads. Thus this solution leads to higher latency, more power consumption and more information processing per node. On the other hand, to maximize the resource utilization, we can choose to have the minimum number of cluster heads to cover the whole geographical area over which the nodes are distributed. The whole area can be split up into zones, the size of which can be determined by the transmission range of the nodes. This can put a lower bound on the number of cluster heads required. Ideally, to reach this lower bound, a uniform distribution of the nodes is necessary over the entire area. Also, the total number of nodes per unit area should be restricted so that the cluster head in a zone can handle all the nodes therein. However, the zone based clustering is not a viable solution due to the following reasons. The cluster heads would typically be centrally located in the zone, and if they move, new cluster heads have to be selected. It might so happen that none of the other nodes in that zone are centrally located. Therefore, to find a new node which can act as a cluster head with the other nodes within its transmission range might be difficult. Another problem arises due to non-uniform distribution of the nodes over the whole area. If a certain zone becomes densely populated then the cluster head might not be able to handle all the traffic generated by the nodes because there is an inherent limitation on the number of nodes a cluster head can handle. We propose to select the minimum number of cluster heads which can support all the nodes in the system satisfying the above constraints.

III. CLUSTER HEAD ELECTION PROCEDURE

The network formed by the nodes and the links can be represented by an undirected graph *G=(V,E)* where *V* represents the set of nodes *vi* and *E* represents the set of links e_i . Dominant set S is subset of V(G).such that

Union of N(V)=V(G)

Here *N(V)* is the *neighborhood* of node *v* , defined as

$$d_v = |N(v)| = \sum_{v' \in V, v' \neq v} \{dist(v, v') < tx_{range}\}$$

where *tx range* is the transmission range of *v* .

Clustering Algorithm use a *combined weight* metric to search dominant set, the combined weight is composed by cluster head degree, battery power, mobility, distance. The *Cluster head election procedure* consists of eight steps as described below:

Step 1**.** Find the neighbors of each node *v* which defines its *degree——dv* as

$$d_v = |N(v)| = \sum_{v' \in V, v' \neq v} \{dist(v, v') < tx_{range}\}$$

Step2: Compute the *degree-difference*
$\Delta_v = |d_v - \delta|$
for every node *v* . Here δ is ideal node number of a cluster except the cluster head.

Step3: For every node, compute the *sum of the distances*, *D v* ,with all its neighbors, as

$$D_v = \sum_{v' \in N(v)} \{dist(v, v')\}$$

Step 4. Compute the running average of the speed for every node till current time*T* . This gives a measure of mobility and is denoted by *M v* , as

$$M_v = \frac{1}{T} \sum_{t=1}^{T} \sqrt{(X_t - X_{t-1})^2 + (Y_t - Y_{t-1})^2}$$

Where (X_t, Y_t) and (X_{t-1}, Y_{t-1}) are the coordinates of the node *v* at time *t* and t-1 respectively.

Step 5. Compute the cumulative time, *P v* during which a node *v* acts as a cluster head. *P v* implies how much battery power has been consumed which is assumed more for a cluster head than an ordinary node.

Step 6. Calculate the *combined weight W v* for each node *v,*

$$W_v = w_1 \Delta v + w_2 Dv + w_3 Mv + w_4 Pv$$

w1, w2, w3, w4 are the *weighing factors* for the corresponding system parameters and

w1+ w2+ w3+ w4=1.

Step 7. Choose that node with the smallest *W v* as the cluster head. All the neighbors of the chosen cluster head are no longer allowed to participate in the election procedure.

Step 8. Repeat steps 2---7 for the remaining nodes not yet selected as a cluster head or assigned to a cluster.

IV. PROPOSED WORK

Factors that influence the implementing the GA

A brief discussion of four factors is given below:

1. *degree-difference*: $\Delta_v = |d_v - \delta|$ for every node *v* . Here δ is ideal node number of a cluster except the cluster head.

2. Battery power (P_v): Obviously, the higher the battery power, the higher the probability that the node will become CH.

3. Degree of mobility: The mobility of the node has great impact on the network lifetime. The topology of the network will be change very frequently due to the high mobility of nodes, which leads to reselection of CHs rapidly.

4. *sum of the distances*, *D v* with all its neighbors, as

$$D_v = \sum_{v' \in N(v)} \{dist(v, v')\}$$

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

Optimization Approach For Cluster Head Selection Using GA:

Algorithm:

```
Alg. Clustering_GA(int chromosome[][] )
  {
   Take dataset(chromosome matrix) according to the node's
neighbourhood at time t;
while(not end of all chromosome in chromosome matrix)
  {
   Take the first row(chromosome) from chromosome
matrix;
   Generate the Gene matrix using the parameter Δv, Dv, Mv
  Pv from the first chromosome row;
   while(convergence criteria is not met )
   {
    Calculate the Wv , value for each Gene (For i=1 to 4)
    {  Wvi = w1Δv + w2Dv + w3Mv + w4Pv
       Wv , = Wv + Wvi
        If(i==4)
        { j=1;
          b[j]= Wv
          j++;
         }
    }
     Maximum and Minimum value is taken from b array;
     Minimum value of b array position row is replaced
     Maximum value of b array position row;
    Getting a new Gene matrix ;
    Take two parent from Gene matrix;
     Mod_Gene[][]=Crossover(Gene);
     Mutation(Mod_Gene[][]);
  }/End For/
  }/End While/
 One of the CH is choosen from the chromosome;
 Take another chromosome;
 }/End main while/
 A set of CH will be choosen among the data set;
 The duplicate node in the set will be deleted to get the
 desired result;
}/End of alg./
```

III. METHODOLOGY

Our goal is to search best nodes among hundreds of nodes, so that they can act as CHs.

Conventional search methods are not robust, while the GA is a search procedure that uses random choice as a tool to guide a highly exploitative search through a coding of a parameter space. According to Goldberg the GA has 4 major characteristics:

1. GAs with a coding of the parameter set, not the parameters themselves.
2. GAs search from a population of points, not a single point.
3. GAs use payoff (objective function) information, not derivatives or other auxiliary knowledge.
4. GAs use probabilistic transition rules, not deterministic rules.

In many optimization methods, we move carefully from a single point in the decision space to the next using some transition rule to determine the next point. This point-to-point method is dangerous because it is a perfect prescription for locating false peaks in multi modal (many peaked) search spaces. By contrast, GA works from a rich database of points simultaneously (a population of strings), climbing many peaks in parallel; thus, the probability of finding a false peak is reduced. A GA starts with a population of strings and thereafter generates successive populations of strings. A simple GA consists of three operators:

1. Reproduction
2. Crossover
3. Mutation

The chromosome of the GA contains all the building blocks to a solution of the problem at hand in a form(fig-1) that is suitable for the genetic operators and the fitness function. Each individual node is represented by a 4 number called `gene'. These four parameter which define the feature of the node and are represented as follows:

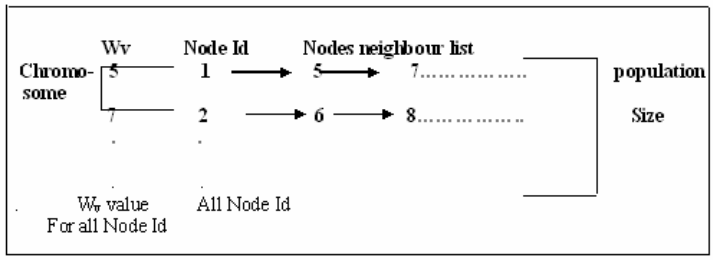
Node ID ⟶ X1 X2 X3 X4

X1: degree-difference
X2: Battery power (Pv),
X3: its degree of mobility, and
X4: sum of the distances

Let's take an example. To start off, select an initial chromosome of total population are neighbours of particular node ID . Here, we select a population of size equal to the no of nodes . Then we have to operate on each chromosome using the 4 parameter for each neighbor nodes of particular node ID. Corresponding node ID has a cluster haead that sould be determined by some fitness value. This value can be evaluated from a fitness function,

$$f(x) = f(x1; x2; x3; x4) = W_1*v + W_2*P_v + W_3*M_v + W_4*D_v.$$

case of Ad-hoc the fitness function depends upon the four factors, discussed in above. And minimum of f(x) should be selected as cluster head. A generation of the GA begins with reproduction. We select the mating pool of the next generation by spinning the weighted roulette wheel four times. From this, the best string get more copies, the average stay even, and the worst die off. Above procedure should be applied for each of the chromosome.



Figure(4.1)(Array Representation of Nodes)

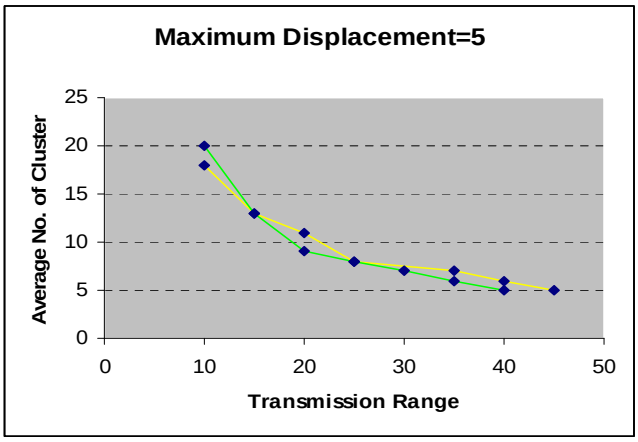
IV. GRAPHICAL ANALYSIS

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

Here, we have shown the comparison between deterministic approach and GA-based approach of weighted clustering algorithm. And we see that sometime genetic algorithm based approach is better than the deterministic approach which is shown in figure(6.5).and sometime show both approach produces the same number of clusterheads as well as cluster. Sometime deterministic gives the lower number of cluster than the number of cluster in GA-based approach. In figure(6.5) green color curve represents the deterministic approach of clustering and yellow color curve represents the GA-based approach .How average number of cluster are changing with respect to the varying transmission range with fixed displacement equal to 5

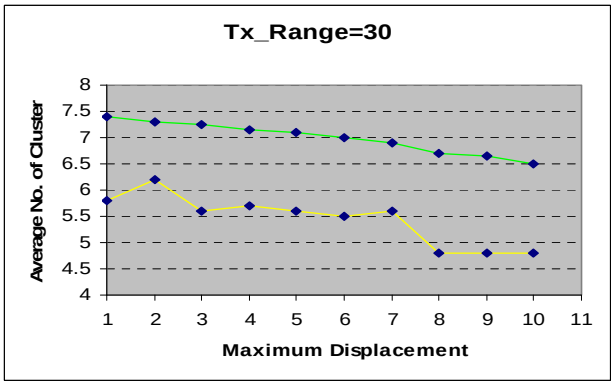
In figure (6.6) shows the comparison of deterministic and GA-based approach between average number of cluster and varying displacement. and we see that GA-based approach always provides the better result than the deterministic approach.

In figure(6.7) shows the comparison of deterministic and GA-based approach between Connectivity and Transmission range .Here connectivity can be measured by the probability that a node is reachable to any other node. For a single component graph ,any node is reachable to the any other node and the connectivity is 1.If the network does not not result in single component graph, then we can say that all the other node in the largest component can communicate with each other and the connectivity can be ratio of the cardinality of the largest component to the cardinality of the graph. From figure(6.7) we have shown the transmission range of the cluster head can be large enough to yield the connected network. If we compare the deterministic approach and GA-based approach ,there we have shown GA gives the better connectivity than the deterministic approach. A well connected graph can be obtained at the cost of a higher transmission range. If we see the graph of transmission range versus average number of cluster heads. There we can see the cluster head will be minimum by incrementing the transmission range .But in GA-based approach gives the better result than deterministic approach. So that in respect of connectivity ,GA-based approach gives the better result.



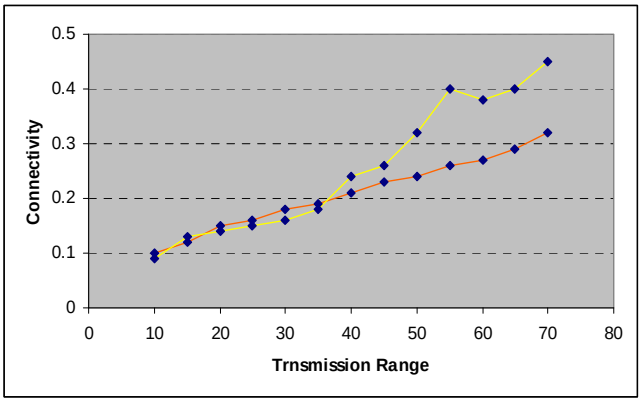
Green Curve = Deterministic Approach
Yellow Curve = GA-based Approach

Figure(6.5)(Comparison Between Deterministic and Soft Computing Approach with Fixed Displacement)



Green Curve = Deterministic Approach.
Yellow Curve = GA-based Approach

Figure(6.6)Comparison Between Deterministic and Soft Computing Approach With Fixed Transmission Range)



Yellow Color Curve= GA-Based Approach
Red Curve = Deterrministic Approach
Figure(6.7) Connectivity Vs Transmission Range

V. CONCLUSION

From the graphical analysis, we have done comparison analysis between deterministic WCA and GA-based WCA and there we have seen that, we can not get always optimistic result in genetic algorithm because genetic algorithm is a randomized searching technique. We have seen when transmission range increases then average number of clusters decreases (Figure(6.5)),so that connectivity of network should be better to compare with the deterministic WCA.

VI. REFERENCES

[1] D.E. Goldberg, "*Genetic Algorithms in Search, Optimization,and Machine Learning*", Addison-Wesley, 1953.
[2] L. Davis, "Applying Adaptive Algorithms to Epistatic Domains", *Proceedings of International Joint Conference on Artificial Intelligence*, 1985.
[3] Jian Zhang, Bin Wang, Fei Zhang.*School of Computer, WuHan University, WuHan 430072, China."* A Distributed Approach of WCA in Ad-hoc Network"
[4] D. Turgut, B. Turgut, R. Elmasri, & T.V. Le,Optimizing clustering algorithm in mobile ad hoc networks using simulated annealing, *Proc. IEEE Wireless Communication and Networking Conference* .
[5] D.E. Goldberg, "International Conference on Genetic Algorithms", *Proceedings of the Fourth International Conference on Genetic Algorithms*, San Diego, July 13-16, 1991.

VII. AUTHORS' PROFILE

Bhaskar Nandi is a lecturer with the department of Computer Science and Engineering, Seacom Engineering College,Howrah,Kolkata,West Bengal, India. He has a teaching experience of about two years, and 1 year of research experience, ,more than two years of industry experience. His research interest are in soft computing, Ad-hoc Networking, Information Security and Data Mining. He has publication in different national journal and conferences. Presently he is working Data Mining and Network Security.

Subhabrata Barman is a Senior lecturer with the department of Computer Science and Engineering, Haldia Institute of Technology, Haldia, West Bengal, India. He has a teaching experience of more than 6 years and a research experience of more than 2 years. His research interests are in the field of Mobile Networking and Computing, Computational Intelligence, Image Processing, Speech and Signal Processing. He has several publications in several national and international conferences and journals. Currently he is working in the area QoS issues and Energy Management in Wireless Adhoc and Sensor Networks.

Soumen Paul is an Assistant Professor with department of the Information Technology, Haldia Institute of Technology, Haldia, West Bengal, India. He has a teaching experience of more than 8 years, industry experience of 11 years and a research experience of more than 2 years. His research interests are in the field of Control Engineering, Soft Computing and Mobile Networking. He has publications in several national and international conferences and journals. His doctoral work is in the area of Deadbeat realization of linear, non-linear, time invariant control systems of nth order.

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

Multi-Product Inventory Optimization using Uniform Crossover Genetic Algorithm

S.Narmadha
Assistant Professor
Department of Computer Science and Engineering
Park College of Engineering and Tekhnology
Coimbatore – 641659, Tamilnadu, India

Dr.V.Selladurai
Professor and Head
Department of Mechanical Engineering
Coimbatore Institute of Technology
Coimbatore – 641014, Tamilnadu, India

G.Sathish
Research Scholar
Department of Computer Science and Engineering
Anna University – Coimbatore, Tamilnadu, India

***Abstract* - Inventory management is considered to be an important field in Supply Chain Management because the cost of inventories in a supply chain accounts for about 30% of the value of the product. The service provided to the customer eventually gets enhanced once the efficient and effective management of inventory is carried out all through the supply chain. The precise estimation of optimal inventory is essential since shortage of inventory yields to lost sales, while excess of inventory may result in pointless storage costs. Thus the determination of the inventory to be held at various levels in a supply chain becomes inevitable so as to ensure minimal cost for the supply chain. The minimization of the total supply chain cost can only be achieved when optimization of the base stock level is carried out at each member of the supply chain. This paper deals with the problem of determination of base-stock levels in a ten member serial supply chain with multiple products produced by factories using Uniform Crossover Genetic Algorithms. The complexity of the problem increases when more distribution centers and agents and multiple products were involved. These considerations leading to very complex inventory management process has been resolved in this work.**

Keywords: Supply Chain Management, Inventory Optimization, Base Stock, Uniform Crossover, Genetic Algorithm (GA), Supply Chain Cost

I. INTRODUCTION

Supply Chain Management (SCM) is an efficient management of the complete end to end process, starting from the design of the product or service to the time when it has been sold, consumed and finally gotten rid of by the consumer. This complete process includes product design, procurement, planning and forecasting, production, distribution, fulfillment and after sales supports. A company's competitiveness in the global economy can be increased only with the aid of effective SCM. This involves complex strategic, tactical and operational decisions that often require an in-depth understanding of industry-specific issues, which ranges from network design to production sourcing and from production planning and inventory management to scheduling [1].

The inventory management problem is one of maintaining an adequate supply of some item to meet an expected pattern of demand, while striking a reasonable balance between the cost of holding the items in inventory and the penalty (loss of sales and goodwill, say) of running out. The item may be a commodity sold by a store; it may be spare machine parts in a factory; it may be railway wagons; it may be cash in the bank to meet the customers' demand. It is indeed surprising to find that a very wide variety of seemingly different problems can be mathematically formulated as an inventory-control problem. There are, of course, several different models of inventory systems. There are three types of expenses associated with inventory systems. The relative importance of these will depend on the specific system. They are: (i) administrative cost of placing an order, called reorder cost or set cost; (ii) cost of maintaining an inventory, called inventory holding cost a carrying cost, which includes storage charge, interest, insurance, etc., a (iii) shortage cost is a loss of profit, goodwill, etc., when run out of stock. All the above should be optimized for efficient supply chain management.

A. Inventory Control in Supply Chain Management

It has been stated by several people that the focus point of supply chain management is inventories and inventory control. To transfer their focus from scheming logistical costs to investigate supply chains [2] few food manufacturers and grocers formed Efficient Consumer Response in the year 1992. The major competitive factor for companies focused on value creation for end consumers is the customer service. In general, firms hold inventory for two major reasons, to lessen costs and to improve customer service. The inspiration for each varies as firms stabilize the problem of having too much inventory (which can direct to high costs) versus having very small inventory (which can direct to lost sales) [3].

Supply chain management leads to cost savings, mainly in the course of lessening in inventory. Inventory costs have got reduced by about 60% from 1982, whereas transportation costs have fallen by 20% [4]. These cost savings have led many people to follow inventory-reduction strategies in the supply chain. To deal with inventory, firms make use of one of three common approaches. First of all, the majority of retailers make use of an inventory control approach, monitoring inventory levels by item. The second thing is, manufacturers are typically more concerned with production scheduling and use flow management to deal with inventories.

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

Third, numerous firms (for the majority part those handling raw materials or in extractive industries) do not keenly deal with inventory [5].

The inventory management is influenced by the nature of demand, depending on whether demand is derived or independent. Independent demand comes up from demand for an end product. End products are found all through the supply chain. By definition, a self-governing demand is uncertain, meaning that extra units or safety stock must be accepted to guard against stock outs. While managing uncertainty, the objective should be to minimize the inventory levels and also meet customer expectation. Supply chain coordination can reduce the ambiguity of intermediate product demand, in that way reducing inventory costs [3, 6].

Since Ford Harris' renowned Economic Order Quantity (EOQ) model was first proposed in 1913, the inventory control has been rewarded immense awareness for a long time because of its significance in the cost control. To lessen the total expected inventory costs per unit time while satisfying the customer demand on time [7] is one of the major objectives. Inventory control for large-scale supply chains is well recognized [8-10] as an essential problem with several applications together with manufacturing systems, logistics systems, communication networks, and transportation systems [11]. It is essential to locate the apt mechanism for coordinating the inventory processes that are controlled by independent partners, in order to find out the right ordering quantity and inventory level amid partners in the chain. For example, the manufacturer make use of the periodic review and lot sizing policy to manage its inventory and the retailer employs the periodic review with target stock level to control its inventory and more [12].

B. Inventory Optimization in Supply Chain Management

The effective management of the supply chain has become unavoidable these days due to high expectation in customer service levels [13]. The supply chain cost was immensely influenced by the overload or shortage of inventories. Thus inventory optimization has transpired into one of the most important topics as far as supply chain management is considered [14-16].

To exploit economies of scale and order in large lots, the important issues in supply chain is to optimize the inventory level by considering various costs in maintaining a high service level towards the customer. Since, the cost of capital tied up in inventory is more, the inventory decision in the supply chain should be coordinated without disturbing the service level. The coordination of inventory decision within an entity is viable, but not between the entities. So the integration of the entities to centralize the inventory control is needed.

Inventory Optimization [IO] application organizes the latest techniques and technologies, thereby assisting the improved inventory visibility, the enhancement of inventory control and its management across an extended supply network. Some of the design objectives of inventory optimization are to optimize inventory strategies, thereby enhancing customer service, reducing lead times and costs and meeting market demand [14-16]. The design and management of the storage policies and procedures for raw materials, work-in-process inventories, and typically, final products are illustrated by the inventory control. The costs and lead times can be reduced and the responsiveness to the changing customer demands can be significantly improved and subsequently inventory can be optimized by the effective handling of the supply chain [17].

There are several reasons for manufacturers' increasing focus on optimizing inventory by applying the latest tools and techniques for inventory control. Traditionally, competitive pressure has always driven manufacturers to seek enhanced capabilities to reduce inventory levels; to enhance service levels and supply availability; and to establish the right product inventory mix and level in each geography and channel. A key driver of the renewed focus on inventory lies in the recognition that traditional techniques are failing to reign in inventories in the wake of increased supply chain complexity. This complexity is characterized by increased uncertainty. Demand is more volatile and therefore less predictable. This is true not only for aggregate demand but for forecasting splits and volumes across channels and markets. Traditionally three strategies have been employed by manufacturers to address uncertainty; a) increase inventory levels to hedge against uncertainty; b) develop supply chain flexibility to be more responsive to uncertainty; c) improve forecast accuracy so that less uncertainty propagates to the manufacturing floor. Inventory optimization techniques and technologies map to the flexibility and accuracy strategies. [18].

Inventory Optimization characterizes the supply network uncertainty present in a variety of specific steps or links in manufacturing and distribution processes. Advanced mathematical models are then solved to identify optimal inventory policies, stocking locations, or quantities. The uncertainty addressed by IO include: demand uncertainty, cycle time variability and replenishment lead time variability.[18] Efficient management of the supply chain, i.e. the reduction of the costs and lead times and vastly enhanced responsiveness to the changing customer demands lead to an optimized inventory.

II. RELATED WORKS - REVIEW

A. Review of Base-Stock based Inventory Control Models

The inventory control problem for a single class assembly network which operates under a modified echelon base-stock policy was studied by Vecchio and Paschalidis [19]. An approach to find close-to-optimal echelon stock levels that minimize inventory costs while guaranteeing stockout probabilities to stay below some predefined levels was developed by them. They reduced the safety stock selection to a deterministic nonlinear optimization problem on the basis of the large deviations techniques. In addition, they analyzed as to how a supplier can interact with a buyer to reach a mutually beneficial mode of operations, using their inventory control approach. Their interaction takes the form of a supply contract by which explicit QoS guarantees is enforced. The applications in a distributed fashion with neither the

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

supplier nor the buyer revealing their corresponding cost structures applying the joint optimization algorithm was proposed by the authors.

A model of supply chain consisting of n production facilities in tandem and producing a single product class was considered by Ioannis CH. Paschalidis et al. [20]. The finished goods inventory maintained in front of the most downstream facility is used to meet the external demand while backlogging of unsatisfied demand was performed. The facility at stage 1 produced if inventory has fallen below a certain level w_i and idles otherwise on the basis of a base-stock production policy adopted at each stage of the supply chain. In order to minimize expected inventory costs at all stages subject to maintaining the stock out probability at stage1 below a prescribed level, they necessitated the optimization of the hedging vector $W= (w_1,...,w_n)$. They made assumptions on demand and production processes that included auto correlated stochastic processes, which were relatively general modeling. They have combined analytical (Large derivations) and sample path based (perturbation analysis) techniques to solve the stochastic optimization problem. The existence of a natural synergy between those two approaches has been demonstrated.

An attempt was made to optimize the inventory (i.e. base-stock) levels of a single product at different members in a serial supply chain with the objective of minimizing the Total Supply Chain Cost (TSCC), by Sudhir Ryan Daniel and Chandrasekharan Rajendran [21] and P. Radhakrishnan et al. [22]. The performance measure considered, which is a good representation of the system-wide total cost is the TSCC. In order to optimize the base-stock levels, a genetic algorithm (GA) has been proposed. To analyze the performance of the supply chain (operating with deterministic and stochastic replenishment lead times) for the base-stock levels that are generated by the proposed GA and other solution procedures considered in this study, different supply chain settings are simulated. They demonstrated that their proposed GA required significantly less computing effort to perform very well in terms of yielding solutions that are not significantly different from the optimal solutions (obtained through complete enumeration of solution space).

A beneficial industry case applying Genetic Algorithms (GA) has been proposed by K.Wang and Y.Wang [23]. The case has made use of GAs for the optimization of the total cost of multiple sourcing supply chain system. The system has been exemplified by a multiple sourcing model with stochastic demand. A mathematical model has been implemented to portray the stochastic inventory with the many to many demand and transportation parameters as well as price uncertainty factors A genetic algorithm which has been approved by Lo [24] deals with the production-inventory problem with backlog in the real situations, with time-varied demand and imperfect production due to the defects in production disruption with exponential distribution. Besides optimizing the number of production cycles to generate a (R, Q) inventory policy, an aggregative production plan can also be produced to minimize the total inventory cost on the basis of reproduction interval searching in a given time horizon. The inventory levels across supply chain members were obtained with the aid of a search routine.

B. Review of Optimization based Inventory Control Models

Sukran Kadipasaoglu et al. [1] provided a study on the market characteristics and competitive priorities, manufacturing environment, logistics and distribution activities, and supply chain planning and control activities for polymer manufacturers. They have described polymer distribution network optimization, production/distribution planning, production scheduling, demand management, available-to-promise, and inventory planning activities pertaining to supply chain planning and control. Besides, they illustrated the applications existing in a commercial DSS that support these activities. They have as well described about diverse issues that continue to confront supply chain managers in polymer manufacturing. It encompasses forecasting for the huge number of product–customer combinations, identification of safety stock requirements, administering production schedule changes, business process management throughout DSS implementation and data mapping for decision support. Their research contributes to the supply chain literature by proffering a suitable context for investigating supply chain-related issues. Through discussion and characterization of the polymer supply chain, they recognized the specific issues of concern to potential researchers and to supply chain professionals.

The effect of product variety on supply-chain performance, which is measured in terms of expected lead time and expected cost at the retailers, was analyzed by Ulrich W. Thonemann and James R. Bradley [25]. They took a supply chain with a single manufacturer and multiple retailers into account. If setup times are significant, the effect of product variety on cost where the cost increases proportionally to the square root of product variety is substantially greater than that suggested by the risk-pooling literature for perfectly flexible manufacturing processes. An illustration that underestimates the cost of product variety, leads companies to offer product variety that is greater than optimal was made as well. In conclusion, they illustrated that by reducing the setup time, the unit manufacturing time, the number of retailers, or the demand rate the supply-chain performance can be managed. The fact that complex mathematical approaches are often not applied in practice was recognized by the authors. Nevertheless, practitioners who used the simple models to estimate the effect of their decisions often appreciated these models.

The inventory and supply chain managers are mainly concerned holding of the excess stock levels and hence the increase in the holding cost. Meanwhile, there is possibility for the shortage of products. For the shortage of each product there will be a shortage cost. Holding excess stock levels as well as the occurrence of shortage for products lead to the increase in the supply chain cost. The factory may manufacture any number of products, each supply chain member may consume a few or all the products and each product is manufactured using a number of raw materials sourced from many suppliers. All these factors pose additional

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

challenge in extracting the exact product and the stock levels that influence the supply chain cost heavily.

Many well-known algorithmic advances in optimization have been made, but it turns out that most have not had the expected impact on the decisions for designing and optimizing supply chain related problems. For example, some optimization techniques are of little use because they are not well suited to solve complex real logistics problems in the short time needed to make decisions. Also some techniques are highly problem-dependent and need high expertise. This adds difficulties in the implementations of the decision support systems which contradicts the tendency to fast implementation in a rapidly changing world. IO techniques need to determine a globally optimal placement of inventory, considering its cost at each stage in the supply chain and all the service level targets and replenishment lead times that constraint each inventory location about the estimation of the exact amount of inventory at each point in the supply chain free of excesses and shortages although the total supply chain cost is minimized. Owing to the fact that shortage of inventory yields to lost sales, whereas excess of inventory may result in pointless storage costs, the precise estimation of optimal inventory is indispensable [26]. In other words, there is a cost involved in manufacturing any product in the factory as well as in holding any product in the distribution center and agent shop. More the products manufactured or held, higher will be the holding cost. Along with this, low lead time results in

III. OBJECTIVES

The supply chain cost can be minimized by maintaining optimal stock levels in each supply chain member. There is a necessity of determining the inventory to be held at different stages in a supply chain that will minimize the total supply chain cost i.e., minimizing holding and shortage cost**.** The inventory control for more number of products along with different levels of supply chain is a complex task. The approach aims to make use of the meta heuristic algorithms like Genetic algorithm for the prediction of the optimal stock levels to be maintained, so as to minimize the total supply chain inventory cost, comprising holding and shortage costs at all members of the supply chain. The genetic algorithm is proposed that considers all these factors that are mentioned hitherto such that the analysis paves the way for minimizing the supply chain cost by maintaining optimal stock levels in each supply chain member.

A. Genetic Algorithm

Genetic algorithm is a randomized search methodology having its roots in the natural selection process. Initially the neighborhood search operators (crossover and mutation) are applied to the preliminary set of solutions to acquire generation of new solutions. Solutions are chosen randomly from the existing set of solutions where the selection probability and the solution's objective function value are proportional to each other and eventually the aforesaid operators are applied on the chosen solutions. Genetic algorithms have aided in the successful implementation of solutions for a wide variety of combinatorial problems.

The robustness of the Genetic algorithms as search techniques have been theoretically and empirically proved [27]. The artificial individual is the basic element of a GA. An artificial individual consists of a chromosome and a fitness value, similar to a natural individual. The individual's likelihood for survival and mating is determined by the fitness function [28]. In accordance with the Darwin's principle, individuals superior to their competitors, are more likely to promote their genes to the next generations. In accordance with this concept, in Genetic Algorithms, a set of encoded parameters are mapped into a potential solution, named chromosome, to the optimization problem [29]. The population of candidate solutions is obtained through the process of selection, recombination, and mutation performed in an iterative manner. [30].

Chromosomes refer to the random population of encoded candidate solutions with which the Genetic algorithms initiate with. [27]. Then the set (called a population) of possible solutions (called chromosomes) are generated [31]. A function assigns a degree of fitness to each chromosome in every generation in order to use the best individual during the evolutionary process [32]. In accordance to the objective, the fitness function evaluates the individuals [30]. Each chromosome is evaluated using a fitness function and a fitness value is assigned. Then, three different operators- selection, crossover and mutation- are applied to update the population. A generation refers to an iteration of these three operators [33]. The promising areas of the search space are focused in the selection step. The selection process typically keeps solutions with high fitness values in the population and rejects individuals of low quality [30]. Hence, this provides a means for the chromosomes with better fitness to form the mating pool (MP) [31]. After the process of Selection, the Crossover is performed.

B. Uniform Crossover

In the crossover operation, two new children are formed by exchanging the genetic information between two parent chromosomes. Multipoint crossover defines crossover points as places between loci where an individual can be split. Uniform crossover generalizes this scheme to make every locus a potential crossover point. A crossover mask, the same length as the individual structure is created at random and the parity of the bits in the mask indicate which parent will supply the offspring with which bits. This method is identical to discrete recombination.

Consider the following two individuals with 11 binary variables each:

Individual 1	0 1 1 1 0 0 1 1 0 1 0
Individual 2	1 0 1 0 1 1 0 0 1 0 1

For each variable the parent who contributes its variable to the offspring is chosen randomly with equal probability. Here, the offspring 1 is produced by taking the bit from parent 1 if the corresponding mask bit is 1 or the bit from parent 2 if the corresponding mask bit is 0. Offspring 2 is created using the inverse of the mask, usually.

Sample 1	0 1 1 0 0 0 1 1 0 1 0
Sample 2	1 0 0 1 1 1 0 0 1 0 1

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

After crossover the new individuals are created:

offspring 1 1 1 1 0 1 1 1 1 1 1 1
offspring 2 0 0 1 1 0 0 0 0 0 0 0

Uniform crossover has been claimed to reduce the bias associated with the length of the binary representation used and the particular coding for a given parameter set. This helps to overcome the bias in single-point crossover towards short substrings without requiring precise understanding of the significance of the individual bits in the individual's representation. How uniform crossover may be parameterized by applying a probability to the swapping of bits was demonstrated by William M. Spears et al.[34].

This extra parameter can be used to control the amount of disruption during recombination without introducing a bias towards the length of the representation used. The chromosome cloning takes place when a pair of chromosomes does not crossover, thus creating off springs that are exact copies of each parent [28].

The ultimate step in each generation is the mutation of individuals through the alteration of parts of their genes [26]. Mutation alters a minute portion of a chromosome and thus institutes variability into the population of the subsequent generation [27]. Mutation, a rarity in nature, denotes the alteration in the gene and assists us in avoiding loss of genetic diversity [26]. Its chief intent is to ensure that the search algorithm is not bound on a local optimum [28].

IV. Inventory Optimization Analysis using Uniform Crossover Genetic Algorithm

The proposed method uses the Genetic Algorithm with Uniform Crossover to study the stock level that needs essential inventory control. This is the pre-requisite information that will make any kind of inventory control effective. In practice, the supply chain is of length *n*, means having *n* number of members in supply chain such as factory, distribution centers, suppliers, retailers and so on. The exemplary supply chain taken for the implementation of the proposed method is shown in Fig. 1.

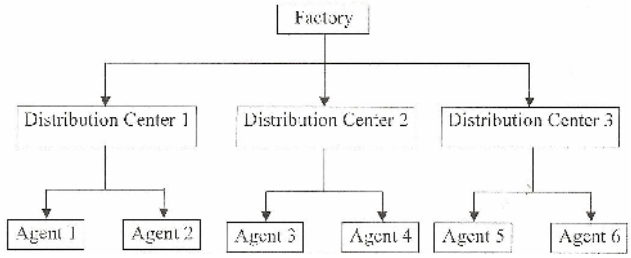


Fig. 1 Three Stage - 10 Member Supply Chain

Fig. 1 exhibits that a factory is the parent of the chain and it is having three distribution centers Distribution Center 1, Distribution Center 2 and Distribution Center 3. Each distribution center further comprises of several agents but as stated in the example case, each Distribution center is having two agents. So, in aggregate there are six agents, Agent 1 and Agent 2 for Distribution Center 1, Agent 3 and Agent 4 for Distribution Center 2 and Agent 5 and Agent 6 for Distribution Center 3. The factory manufactures different products that would be supplied to the distribution centers. From the distribution center, the stocks will be moved to the corresponding agents.

To make the inventory control effective, the most primary objective is to predict where, why and how much of the control is required which is made through the proposed GAUX methodology. The proposed methodology is aimed at determining the specific product that needs to be concentrated on and the amount of stock levels of the product to be maintained by the different members of the supply chain. The methodology also analyses whether the stock level of the particular product needs to be in abundance, in order to avoid shortage of the product or needs to be held minimal in order to minimize the holding cost.

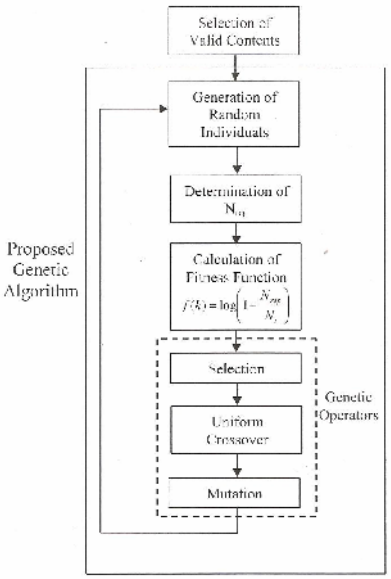


Fig. 2 Genetic Algorithm steps for the proposed inventory management analysis

The methodology as shown in Fig. 2 would analyze the past records very effectively and thus facilitate efficient inventory management with the aid of Genetic Algorithm. The analysis is initiated by the selection of valid records. The validation of records is done over the records of past periods. The stock levels at the different supply chain members are held in the dataset for different products, namely P1, P2, P3, P4, P5, P6, P7, etc. Seven products have been considered for the analysis, hence the stock levels for the seven products at each member of the chain throughout the past period are considered as data set as shown in the Table 1. For the valid record set selection, records having nil values are neglected and the records having positive or negative values are selected for the analysis. This can be done by means of clustering algorithms, extraction algorithms or by any of the data mining functions. Hence the extraction function results in data sets having either positive or negative values. The record set having positive values represents excess stock levels and the negative values represent shortage level of a particular product at a particular member of the supply chain. Then the

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

data set is subjected to Genetic Algorithm and the various steps performed in the genetic algorithm are discussed below.

A. Generation of Individuals

The randomly generated initial chromosome is created by having the stock levels within the lower limit and the upper limit for all the contributors of the supply chain, factory and the distribution centers. As known, chromosome is constituted by genes which defines the length of the chromosomes. The stock level of each member of the chromosome is referred as gene of the chromosome. Hence for n length supply chain, the chromosome length is also n . Since a 10 member supply chain is used for illustration, the length of the chromosome n is 10, i.e. 10 genes. And the chromosome representation is pictured in Fig. 3. Each gene of the chromosome is representing the amount of stock that is in excess or in shortage at the respective members of the supply chain.

P3	7000	-200	-600	-500	450	-350	800	-400	700	-600

P2	5000	400	-800	500	445	315	-820	405	-150	100

Fig. 3 Random individual generated for the genetic operation

These kinds of chromosomes are generated for the genetic operation. Initially, only two chromosomes will be generated and from the next generation a single random chromosome value will be generated. The chromosomes thus generated is then applied to find its number of occurrences in the database content by using a Select count ( ) function. The function will give the number of occurrences/ repetitions of the particular amount of stock level for the ten members N_{rep} that are going to be used further in the fitness function.

B. Evaluation of Fitness function

A specific kind of objective function that enumerates the optimality of a solution in a genetic algorithm in order to rank certain chromosome against all the other chromosomes is known as Fitness function. Optimal chromosomes, or at least chromosomes which are near optimal, are permitted to breed and merge their datasets through one of the several techniques available in order to produce a new generation that will be better than the ones considered so far.

The fitness function is given by:

$$f(k) = \log\left(1 - \frac{N_{rep}}{N_t}\right), \quad k = 1,2,3\cdots\cdots, m \qquad (1)$$

where,

N_{rep} is the number of repetitions of records of similar stock levels that occurs throughout the period;

N_t is the total number of records of inventory values obtained after clustering;

m is the total number of chromosomes for which the fitness function is calculated.

In the fitness function, the ratio (N_{rep} / N_t) plays the role of finding the probability of occurrence of a particular record of inventory values; and log [1- (N_{rep} / N_t)] will ensure minimum value corresponding to the maximum probability; So, the fitness function is structured to retain the minimum value corresponding to the various chromosomes being evaluated iteration after iteration and this in turn ensures that the fitness function evolution is towards optimization.

C. Selection

The selection operation is the initial genetic operation which is responsible for the selection of the fittest chromosome for further genetic operations. The fitness function is carried out for each chromosome and the chromosomes are sorted on the basis of the result of the fitness function and ranked. The chromosome generating value as minimum as possible will be selected by the fitness function and will be subjected further to the genetic operations, crossover and mutation.

D. Uniform Crossover

Among the numerous crossover operators in practice, a uniform crossover is chosen in this proposed method for its advantages over the other forms. Uniform crossover is global and less biased when compared to that of standard and one point crossover. Uniform crossover does not select a set of crossover points. It simply considers each bit position of the two parents, and swaps the two bits with a probability of 50%. With large search spaces, a GA using uniform crossover outperforms a GA using one point crossover, which in turn outperforms a GA using two point crossover [35-36]. From the matting pool, two chromosomes are subjected for the uniform crossover. The chromosomes initially selected and after undergoing uniform crossover operation performed in this analysis is pictured in Fig. 4. As soon as the crossover operation is completed, the genes of the two chromosomes present get interchanged.

Before Crossover

P3	7000	-200	-600	-500	450	-350	800	-400	700	-600

P2	5000	400	-800	500	445	315	-820	405	-150	100

After Crossover

P3	7000	-150	-820	-400	315	-350	-750	405	500	600

P2	5000	450	700	-600	350	-200	800	500	-100	150

Fig. 4 Chromosomes after uniform crossover operation

E. Mutation

The crossover operation is succeeded by the final stage of genetic operation known as Mutation. In the mutation, a new chromosome is obtained. This chromosome is totally new from the parent chromosome. The concept behind this is the child chromosome thus obtained will be fitter than the parent chromosome. The performance of mutation operation is shown in Fig. 5.

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

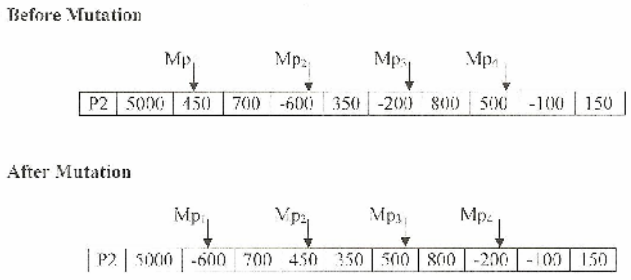


Fig. 5 Chromosome subjected to mutation operation

Four mutation points are chosen as shown in Fig. 5. The mutation is done on the particular gene present at the Mutation points. This pointing of gene is done randomly. Hence, the four mutation points may point any of the ten genes.

The process explained so far will be repeated along with the new chromosome obtained from the previous process. In other words, at the end of each of the iteration, a best chromosome will be obtained. This will be included with the newly generated random chromosome for the next iteration. When the number of iterations is increased then the obtained solution moves very closer to the accurate solution. More the number of iterations the more accurate the optimal solution will be. Eventually with the help of the Genetic algorithm, the best stock level to be maintained in the members of the supply chain could be predicted from the past records, so that the loss due to the holding of excess stock level and shortage level can be reduced leading to an optimal inventory solution.

V. Experimental Results

The approach suggested for the optimization of inventory level and thereby an efficient supply chain management has been implemented in the platform of LabVIEW 2009. The database consists of the records of stock levels held by each member of the supply chain for every period. For implementation, seven different products in circulation with the ten member supply chain network have been considered. A sample set of data from a large database used in the implementation is given in Table 1.

Table 1. A Sample Dataset Constituted by the Product Identification along with its Stock Levels in Each Member of the Supply Chain

PI	F1	DC1	DC2	DC3	A1	A2	A3	A4	A5	A6
7	-371	-736	-299	634	448	756	340	-736	-778	863
5	-407	379	-981	-864	-391	999	-196	307	-171	-529
2	-146	-604	443	746	-561	-734	445	424	-891	-824
4	-962	-524	-685	-254	205	446	-469	108	346	840
3	-834	266	969	965	735	244	-752	133	-554	-939
3	-449	-282	577	-926	-414	-200	-743	850	196	851
4	540	-830	-835	882	-379	768	-635	-112	539	107
3	-778	-313	629	-690	824	-927	850	307	-171	-529
2	351	293	328	-732	357	-566	685	424	-891	-824
1	500	108	490	-345	-236	108	-931	-260	-144	162
5	844	-728	286	740	686	-421	424	-792	-927	-879
4	-321	902	-450	-260	-144	162	238	307	-171	-529
3	775	-394	-520	-792	-927	-879	-507	424	-891	-824
4	794	932	-584	307	-171	-529	-503	108	346	840
2	-122	-686	-620	424	-891	-824	941	133	-554	-939
6	235	464	401	108	346	840	-934	464	401	108
5	218	-848	836	133	-554	-939	-834	-848	836	133
4	489	409	148	850	196	851	-495	-144	162	238
3	-422	638	676	-112	539	107	-440	-927	-879	-507
5	893	520	-423	-736	-778	863	-335	676	-112	539

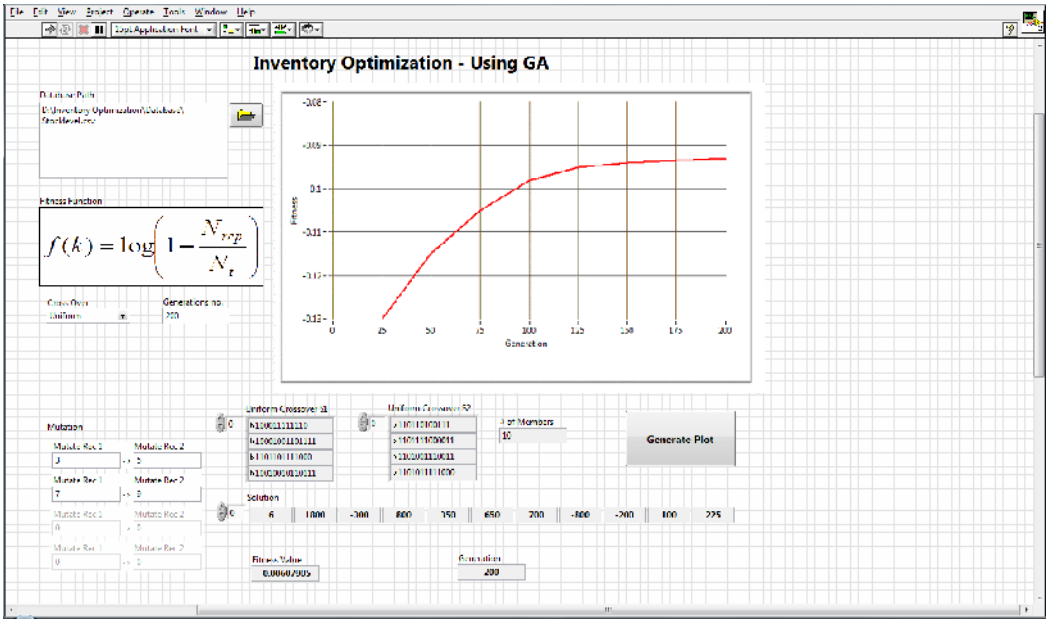


Fig. 6 Inventory Optimization Tool created in LabVIEW 2009

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

In the sample database tabulated in Table 1, the first field comprises of the product Identification (PI) and the other fields are related with the stock levels that were held by the respective ten members of the supply chain network. For example, the first attribute and first field of the database is '7' which refers the Product I.D. '7'. The corresponding fields of the same attribute denote the stock levels of the product I.D. '7' in the respective members of the supply chain. Similarly, different sets of stock levels for different products are held by the database.

As per the proposed analysis based on Uniform Crossover GA, two random initial chromosomes are generated as shown in Fig. 3. These initial chromosomes are subjected for the genetic operators, Uniform Crossover and Mutation. The resultant chromosome thus obtained after the application of crossover and mutation is shown in Fig. 5.

Fig. 6 shows the main window of the tool created for inventory optimization using genetic algorithm in LabVIEW 2009. The tool created is based on the uniform crossover genetic algorithm.

The window displays the fitness function used, uniform crossover sample bit pattern randomly generated with probability of 0.5, the points at which mutation should occur, the end iteration count upon which the fitness function stabilizes, no. of supply chain members and population size.

The best chromosome obtained after the required iterations and the fitness value and plot generated for the iteration value are also displayed.

The organization can decide about the quantum of iterations for running the simulation to arrive at the optimal solution. As long as minimization of the fitness function is still possible, then the iteration continues till such a time that no improvement in the fitness function value is noticeable. After a certain number of iterations, if the fitness function value is not improving from the previous iterations, then this is an indication that the fitness function value is stabilizing and the algorithm has converged towards optimal solution. This inference is useful for deciding the number of iterations for running the GA simulation as well as this may be used as the stopping criteria for the algorithm. For greater accuracy, the number of iterations should be sufficiently increased and run on the most frequently updated large database of past records.

As for our iteration value of '200', the resultant chromosome moved towards the best chromosome after each iterative execution. Hence at the end of the execution of 200^{th} iteration, best chromosome obtained is shown in Fig. 7.

6	1800	-300	800	350	650	700	-800	-200	100	225

Fig. 7 The final best chromosome obtained after 200^{th} iteration

VI. Discussion of Results

The final chromosome obtained from the GA based analysis shown in the Fig. 7 is the inventory level that has the potential to cause maximum increase of supply chain cost. It is inferred that controlling this resultant chromosome is sufficient to reduce the loss either due to the holding of excess stocks or due to the shortage of stocks. By focusing on the excess/shortage inventory levels and initiating appropriate steps to eliminate the same at each member of the chain, it is possible to optimize the inventory levels in the upcoming period and thus minimize the supply chain cost.

The organization should take necessary steps to decrease the production of product 6 in the factory by 1800 units to make up for the predicted excess; increase the inventory level of product 1 by 300 units in distribution center 1 to make up for the predicted shortage, reduce inventory level of product 1 by 800 units and 350 units in distribution centers 2 and 3 respectively to make up for the predicted excess.

Agent 1 should decrease the inventory level of product 6 by 650 units. Agent 2 should decrease the inventory level of product 6 by 700 units. Agent 3 and Agent 4 should increase the inventory level of product 6 by 800 units and 200 units respectively to make up for the predicted excess / shortage. The inventory level of product 6 should be decreased by 100 units and 225 units by Agent 5 and Agent 6 respectively. Thus by following the predicted stock levels, the excess/shortage inventory levels can be avoided in the upcoming period and thus the increase of supply chain cost can also be avoided. The analysis extracts an inventory level that made a remarkable contribution towards the increase of supply chain cost, and in turn enabled to predict the future optimal inventory levels to be maintained in all the supply chain members with the aid of these levels. Therefore it is possible to minimize the supply chain cost by maintaining the optimal stock levels that was predicted from the inventory analysis, and thus making the inventory management more effective and efficient.

VII. Conclusion

Inventory management is an important component of supply chain management. An innovative and efficient methodology that uses Genetic Algorithms with Uniform Crossover to precisely determine the most probable excess stock level and shortage level required for inventory optimization in the supply chain such that the total supply chain cost is minimal is proposed using LabVIEW 2009.

The optimized stock level at all members of the supply chain is obtained by following the proposed genetic algorithm. Thus the proposed work gives a better prediction of stock levels amid diverse stock levels at all members of the supply chain. The complexity of increasing the number of products through the supply chain has been resolved by the proposed approach. Products due to which the members of the supply chain incur extra holding or shortage cost are also determined. More specifically, the inventory is optimized in the whole supply chain regardless of the number of products and the number of members in the supply chain. The proposed approach of inventory management has achieved the objectives which are the minimization of total supply chain cost and the determination of the products due to which the supplier endured either additional holding cost or shortage cost.

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

REFERENCES

[1]. Sukran Kadipasaoglu, Jennifer Captain and Mark James, "Polymer Supply Chain Management", *International Journal on Logistics Systems and Management*, Vol. 4, No. 2, pp. 233-253, 2008.

[2]. R. King and P. Phumpiu, "Reengineering the food supply chain: The ECR initiative in the grocery industry", *American Journal of Agricultural Economics*, Vol. 78, pp. 1181-1186, 1996.

[3]. Frank Dooley, "Logistics, Inventory Control, and Supply Chain Management", CHOICES: The magazine of food, farm and resource Issues, Vol. 20, No. 4, 4th Quarter 2005.

[4]. R. Wilson, 15th Annual State of Logistics Report. Council of Supply Chain Management Professionals, 2004. Available online: http://www.cscmp.org/.

[5]. R. Ballou, "Business logistics/supply chain management", 5th Ed. Upper Saddle River, NJ: Prentice Hall, 2004.

[6]. M. Fisher, "What is the right supply chain for your product?", Harvard Business Review, Mar/Apr., pp. 105-116, 1997.

[7]. Guangyu Xiong and Hannu Koivisto, "Research on Fuzzy Inventory Control under Supply Chain Management Environment", Lecture Notes in Computer Science, Vol. 2658, pp. 673, 2003.

[8]. M.C. Bonney, "Trends in Inventory Management", *International Journal on Production Economics*, Vol. 35, No. 1-3, pp. 107-114, 1994.

[9]. M. Muller, "Essentials of Inventory Management", NewYork: Amer. Manage. Assoc., 2002.

[10]. S. Nahmias, "Production and Operations Analysis", New York: McGraw-Hill/Irwin, 2004.

[11]. Krishnamurthy, Khorrami and Schoenwald, "Decentralized Inventory Control for Large-Scale Reverse Supply Chains: A Computationally Tractable Approach", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 38, No. 4, pp. 551-561, July 2008.

[12]. Kanit Prasertwattana, Yoshiaki Shimizu and Navee Chiadamrong, "Evolutional Optimization on Material Ordering and Inventory Control of Supply Chain through Incentive Scheme", *Journal of Advanced Mechanical Design Systems, and Manufacturing*, Vol. 1, No. 4, pp. 562-573, 2007.

[13]. Mileff, Peter, Nehez, Karoly, "A new inventory control method for supply chain management", 12th International Conference on Machine Design and Production, 2006.

[14]. "Optimization Engine for Inventory Control", Golden Embryo Technologies Pvt. Ltd., Maharastra, India, 2004.

[15]. Jinmei Liu, Hui Gao, Jun Wang, "Air material inventory optimization model based on genetic algorithm", Proceedings of the 3rd World Congress on Intelligent Control and Automation, Vol. 3, pp. 1903 - 1904, 2000.

[16]. C.M. Adams, "Inventory optimization techniques, system vs. item level inventory analysis", 2004 Annual Symposium RAMS - Reliability and Maintainability, pp. 55 - 60, 26-29, Jan, 2004.

[17]. P. Pongcharoen, A. Khadwilard and A. Klakankhai, "Multi-matrix real-coded Genetic Algorithm for minimizing total costs in logistics chain network", *World Academy of Science, Engineering and Technology*, Vol. 26, pp. 458-463, 14-16, December, 2007.

[18]. Greg Scheuffele and Anupam Kulshreshtha, "Inventory Optimization A Necessity Turning to Urgency", SETLabs Briefings, Vol. 5, No. 3, 2007.

[19]. Vecchio and Paschalidis, "Enforcing service-level constraints in supply chains with assembly operations", *Proceedings of IEEE Conference on Decision and Control*, Vol. 5, pp. 5490- 5495, December 2003.

[20]. Ioannis CH. Paschalidis, Yong Liu, Christos G. Cassandras and Christos Panayiotu, "Inventory control for supply chains with service level constraints: A synergy between Large Deviations and Perturbation analysis", *Annals of Operations Research*, Vol. 126, pp. 231-258, 2004.

[21]. J. Sudhir Ryan Daniel and Chandrasekharan Rajendran, "A simulation-based genetic algorithm for inventory optimization in a serial supply chain", *International Transactions in Operational Research*, Vol. 12, pp. 101-127, 2005.

[22]. P. Radhakrishnan, V.M. Prasad and M.R. Gopalan, "Inventory optimization in supply chain management using genetic algorithm", *International Journal of Computer Science and Network Security*, Vol. 9, No. 1, January 2009, pp. 1-8.

[23]. K. Wang and Y. Wang, "Applying Genetic Algorithms to Optimize the Cost of Multiple Sourcing Supply Chain Systems - An Industry Case Study", Studies on Computational Intelligence, Vol. 92, pp. 355-372, 2008.

[24]. Chih-Yao Lo, "Advance of Dynamic Production-Inventory Strategy for Multiple Policies Using Genetic Algorithm", *Information Technology Journal*, Vol. 7, pp. 647-653, 2008.

[25]. Ulrich W. Thonemann and James R. Bradley, "The effect of product variety on supply-chain performance", *European Journal of Operational Research*, Vol. 143 No.3, pp. 548-69, 2002.

[26]. S. Buffett and N. Scott, "An Algorithm for Procurement in Supply Chain Management", *AAMAS-04 Workshop on Trading Agent Design and Analysis*, New York, 2004.

[27]. S. Behzadi, Ali A. Alesheikh and E. Poorazizi, "Developing a Genetic Algorithm to solve Shortest Path Problem on a Raster Data Model", *Journal on Applied Sciences*, Vol. 8, No. 18, pp. 3289-3293, 2008.

[28]. Aphirak Khadwilard and Pupong Pongcharoen, "Application of Genetic Algorithm for Trajectory Planning of Two Degrees of Freedom Robot Arm With Two Dimensions", *Thammasat International Journal on Science and Technology*, Vol. 12, No. 2, April- June 2007.

[29]. M.A. Sharbafi, M. Shakiba Herfeh, Caro Lucas and A. Mohammadi Nejad, "An Innovative Fuzzy Decision Making Based Genetic Algorithm", *World Academy of Science, Engineering and Technology*, Vol. 19, pp. 172-175, May 2006.

[30]. Thomas Butter, Franz Rothlauf, Jorn Grahl, Hildenbrand Jens Arndt, "Developing Genetic Algorithms and Mixed Integer Linear Programs for Finding Optimal Strategies for a Student's Sports Activity", *Working Papers in Information Systems, University of Mannheim*, 2006.

[31]. S.A. Qureshi, S.M. Mirza and M. Arif, "Fitness Function Evaluation for Image Reconstruction using Binary Genetic Algorithm for Parallel Ray Transmission Tomography", *International Conference on Emerging Technologies*, 2006. ICET '06. 13-14, Nov. 2006, pp. 196-201.

[32]. Saifuddin Md. Tareeq, Rubayat Parveen, Liton Jude Rozario and Md. Al-Amin Bhuiyan, "Robust Face detection using Genetic Algorithm", *Journal on Information Technology*, Vol. 6, No. 1, pp. 142-147, 2007.

[33]. M. Soryani and N. Rafat, "Application of Genetic Algorithms to Feature Subset Selection in a Farsi OCR", *World Academy of Science, Engineering and Technology*, Vol. 18, pp. 113-116..

[34]. William M. Spears and K.A. De Jong, "On the Virtues of Uniform Crossover", 4th International Conference on Genetic Algorithms, La Jolla, California, July 1991.

[35]. Syswerda, Gilbert, "Uniform Crossover in Genetic Algorithms", Proc. 3rd International Conference on Genetic Algorithms, Morgan Kaufman Publishing, 1989.

[36]. Riccardo Poli and W.B. Langdon, "On the Search Properties of Different Crossover Operators in Genetic Programming", Proceedings of Genetic Programming'98, Madison, Wisconsin, 1998.

ABOUT AUTHORS

Mrs. S. Narmadha is working as Assistant Professor in the Department of Computer Science and Engineering in Park College of Engineering and Technology, Coimbatore. She obtained her Bachelor's degree in Computer Science and Engineering from Tamilnadu College of Engineering, Coimbatore under Bharathiar University and Master's degree in Mechatronics from Vellore Institute of Technology, Vellore. She is currently pursuing Ph.D. under Anna University, Chennai. She has 8 years of Teaching Experience and 2 years of Industrial experience. She has published 14 papers in International Conferences, 2 papers in International journals and a book on 'Open Source Systems'. She is life member of ISTE. Her field of interest includes Supply Chain Management, Automation, Database Management Systems, Virtual Instrumentation, Soft Computing Techniques and Image Processing.

Dr. V. Selladurai is the Professor and Head, Department of Mechanical Engineering and Principal, Coimbatore Institute of Technology, Coimbatore, India. He holds a Bachelor's degree in Production Engineering, a Master's degree in Industrial Engineering specialisation and a Ph.D. degree in Mechanical Engineering. He has two years of industrial experience and 22 years of teaching experience. He has published over 90 papers in the proceedings of the leading National and International Conferences. He has published over 35 papers in international journals and 22 papers in national journals. His areas of interest include Operation Research, Artificial Intelligence, Optimization Techniques, Non-Traditional Optimization Techniques, Production Engineering, Industrial and Manufacturing Systems, Industrial Dynamics, System Simulation, CAD/CAM, FMS, CIM, Quality Engineering and Team Engineering.

Mr. G. Sathish is a full time research scholar under Anna University, Coimbatore. He holds a Bachelor's degree and a Master's degree in Computer Science and Engineering. He has 8 years of industrial experience. He has published 10 papers in the proceedings of the leading International Conferences. His field of interest includes Supply Chain Management, Optimization Techniques, Data Mining, Knowledge Discovery, Automation, Soft Computing Techniques and Image Processing.

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

Efficient Inventory Optimization of Multi Product, Multiple Suppliers with Lead Time using PSO

S.Narmadha
Assistant Professor
Department of Computer Science and Engineering
Park College of Engineering and Tekhnology
Coimbatore – 641659, Tamilnadu, India

Dr.V.Selladurai
Professor and Head
Department of Mechanical Engineering
Coimbatore Institute of Technology
Coimbatore – 641014, Tamilnadu, India

G.Sathish
Research Scholar
Department of Computer Science and Engineering
Anna University – Coimbatore, Tamilnadu, India

Abstract - With information revolution, increased globalization and competition, supply chain has become longer and more complicated than ever before. These developments bring supply chain management to the forefront of the management's attention. Inventories are very important in a supply chain. The total investment in inventories is enormous, and the management of inventory is crucial to avoid shortages or delivery delays for the customers and serious drain on a company's financial resources. The supply chain cost increases because of the influence of lead times for supplying the stocks as well as the raw materials. Practically, the lead times will not be same through out all the periods. Maintaining abundant stocks in order to avoid the impact of high lead time increases the holding cost. Similarly, maintaining fewer stocks because of ballpark lead time may lead to shortage of stocks. This also happens in the case of lead time involved in supplying raw materials. A better optimization methodology that utilizes the Particle Swarm Optimization algorithm, one of the best optimization algorithms, is proposed to overcome the impasse in maintaining the optimal stock levels in each member of the supply chain. Taking into account the stock levels thus obtained from the proposed methodology, an appropriate stock levels to be maintained in the approaching periods that will minimize the supply chain inventory cost can be arrived at.

Keywords: Supply Chain Management, Inventory Optimization, Base Stock, Multiple Suppliers, Lead Time, Particle Swarm Optimization (PSO), Supply Chain Cost

I. INTRODUCTION

Inventory takes many forms, ranging from raw materials to finished goods. While holding large amounts of inventory enables a company to be responsive to fluctuations in customer demand, the associated costs can be excessive. In order to operate in a lean environment at maximum efficiency levels, companies must minimize all unnecessary expenses, including those associated with production and storage of inventories.

Inventory control is typically a key aspect of almost every manufacturing and/or distribution operation business. The ultimate success of these businesses is often dependent on its ability to provide customers with the right goods, at the right place, at the right time. The right goods are those that the customer wants; the right place is your "available" inventory, not the supplier's warehouse, and in today's economy the right time is immediately.

Failure to have the right goods in the right place at the right time often leads to lost sales and profits and, even worse, to lost customers. Today's reality is that there is very little differentiation between commodity products of the same type, and customers will, more often than not, choose to return to businesses that meet all three conditions, even choosing relatively unknown brands over known brands.

The role of inventory management is to coordinate the actions of all business segments, particularly sales, marketing and production, so that the appropriate level of stock is maintained to satisfy customers' demands. The goal of inventory management is to balance supply and demand as closely as possible in order to keep customers satisfied and drive profits.

Inventory management is a fundamental requisite to supply chain optimization. The processes and controls of effective inventory management are critical to any successful business. Since it is rarely the case that any business has the luxury of unlimited capital, inventory management involves important decisions about what to buy or produce, how much to buy or produce and when to buy or produce within the capital limits. These are "value decisions." Excessive inventory investments can tie up capital that may be put to better use within other areas of the business. On the other hand, insufficient inventory investment can lead to inventory shortages and a failure to satisfy customer demand. A balance must be struck and maintained.

The aim of inventory management is to reduce inventory holdings to the lowest point without negatively impacting availability or customer service levels. This can be done while still maximizing the business' ability to exploit economies of scale to positively impact profitability.

Inventory optimization takes inventory management to the next level, enabling businesses to further reduce inventory levels while improving customer service levels and maximizing capital investments.

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

Inventory management is an ongoing process that relies on inputs from forecasts and product pricing, and should be executable within the cost structure of the business under an overall plan. Inventory control involves three inventory forms of the flow cycle:

- Basic Stock - The exact quantity of an item required to satisfy a demand forecast.
- Seasonal Stock - A quantity buildup in anticipation of predictable increases in demand that occur at certain times in the year.
- Safety Stock - A quantity in addition to basic inventory that serves as a buffer against uncertainty.

The challenge is to weigh the balance in favor of basic stock so that the business holds as little safety stock as possible and provides 'just the right amount' of seasonal stock. However, the predictability of demand has a direct impact on how much safety stock a business must hold. When demand is unpredictable, higher levels of safety stock must be maintained. Therefore, the search for the optimal inventory levels to achieve a lean manufacturing environment becomes a key objective.

A. *Benefits of Inventory Optimization*

The primary function of an Inventory Optimization solution is to allow companies to effectively fulfill demand and identify how to gain additional profits from their inventories. Improved efficiencies through effective resource management and optimization lead to an increase in service level, improved performance against customer request dates and improved return on equity. These gains are derived in three ways: a) System Benefits b) Value-Added Benefits and c) Strategic Benefits.

B. *Particle Swarm Optimization*

In 1995, Kennedy and Eberhartin, inspired by the choreography of a bird flock, first proposed the Particle Swarm Optimization (PSO). In comparison with the evolutionary algorithm, PSO, relatively recently devised population-based stochastic global optimization algorithm, has many similarities and the robust performance of the proposed method over a variety of difficult optimization problems has been proved [1]. In accordance with PSO, either the best local or the best global individual affects the behavior of each individual in order to help it fly through a hyperspace [2]. Simulation of simplified social models has been employed to develop Particle Swarm Optimization techniques. The following are the features of the method [3]:

- The researches on swarms such as fish schooling and bird flocking are the basis of the method.
- The computation time is short and it requires little memory as it is based on a simple concept.
- Nonlinear optimization problems with continuous variables were the initial focus of this method. Nevertheless, problems with discrete variables can be treated by easy expansion of the method. Hence, the mixed integer nonlinear optimization problems with both continuous and discrete variables can be treated with this method.

In addition to PSO, several evolutionary paradigms exist which include Genetic algorithms (GA), Genetic programming (GP), Evolutionary strategies (ES) and Evolutionary programming (EP). Biological evolution is simulated by these approaches which are based on population [4]. Genetic algorithm and PSO are two widely used types of evolutionary computation techniques among the various types of Evolutionary Computing paradigms [5].

PSO and evolutionary computation techniques such as Genetic Algorithms (GA) have many similarities between them. A population of random solutions is used to initialize the system which updates generations to search for optima. Nevertheless, PSO does not have evolution operators such as crossover and mutation that are available in GA.

In PSO, the potential solutions, called particles follow the current optimum particles to fly through the problem space. Every particle represents a candidate solution to the optimization problem. The best position visited by the particle and the position of the best particle in the particle's neighborhood influences its position.

Particles would retain part of their previous state using their memory. The particles would still remember the best positions they ever had even as there are no restrictions for particles to know the positions of other particles in the multidimensional spaces. An initial random velocity and two randomly weighted influences: individuality (the tendency to return to the particle's best previous position), and sociality (the tendency to move towards the neighborhood's best previous position) form each particle's movement [6].

When the neighborhood of a particle is the entire swarm, the global best particle refers to the best position in the neighborhood and in this case, gbest PSO refers the resulting algorithm. Generally, lbest PSO refers the algorithm in cases when smaller neighborhoods are used [5]. A fitness function that is to be optimized evaluates the fitness values of all the particles [6].

PSO uses individual and group experiences to search the optimal solutions. Nevertheless, previous solutions may not provide the solution of the optimization problem. The optimal solution is changed by adjusting certain parameters and putting random variables. The ability of the particles to remember the best position that they have seen is an advantage of PSO [6].

II. RELATED REVIEW

A fresh Genetic Algorithm (GA) approach for the Integrated Inventory Distribution Problem (IIDP) has been projected by Abdelmaguid et al. [7]. They have developed a genetic representation and have utilized a randomized version of a formerly developed construction heuristic in order to produce the initial random population.

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

Pongcharoen et al.[8] have put forth an optimization tool that works on basis of a Multi-matrix Real-coded Generic Algorithm (MRGA) and aids in reduction of total costs associated with in supply chain logistics. They have incorporated procedures that ensure feasible solutions such as the chromosome initialization procedure, crossover and mutation operations. They have evaluated the algorithm with the aid of three sizes of benchmarking dataset of logistic chain network that are conventionally faced by most global manufacturing companies.

A technique to utilize in supply-chain management that supports the decision-making process for purchases of direct goods has been projected by Buffett et al.[9]. RFQs have been constructed on basis of the projections for future prices and demand and the quotes that optimize the level of inventory each day besides minimizing the cost have been accepted. The problem was represented as a Markov Decision Process (MDP) that allows for the calculation of the utility of actions to be based on the utilities of substantial future states. The optimal quote requests and accepts at each state in the MDP were determined with the aid of Dynamic programming. A supply chain management agent comprising of predictive, optimizing and adaptive components called the TacTex-06 has been put forth by Pardoe et al. [10]. TacTex-06 functions by making predictions regarding the future of the economy, such as the prices that will be proffered by component suppliers and the degree of customer demand and then strategizing its future actions so as to ensure maximum profit.

Beamon et al.[11] have presented a study on evaluations of the performance measures employed in supply chain models and have also displayed a framework for the beneficial selection of performance measurement systems for manufacturing supply chains Three kinds of performance measures have been recognized as mandatory constituents in any supply chain performance measurement system. New flexibility measures have also been created for the supply chains. The accomplishment of beam-ACO in supply-chain management has been proposed by Caldeira et al.[12]. Beam-ACO has been used to optimize the supplying and logistic agents of a supply chain. A standard ACO algorithm has aided in the optimization of the distributed system. The application of Beam-ACO has enhanced the local and global results of the supply chain.

A beneficial industry case applying Genetic Algorithms (GA) has been proposed by Wang et al.[13]. The case has made use of GAs for the optimization of the total cost of a multiple sourcing supply chain system. The system has been exemplified by a multiple sourcing model with stochastic demand. A mathematical model has been implemented to portray the stochastic inventory with the many to many demand and transportation parameters as well as price uncertainty factors. A genetic algorithm which has been approved by Lo [14] deals with the production-inventory problem with backlog in the real situations, with time-varied demand and imperfect production due to the defects in production disruption with exponential distribution. Besides optimizing the number of production cycles to generate a (R, Q) inventory policy, an aggregative production plan can also be produced to minimize the total inventory cost on the basis of reproduction interval searching in a given time horizon.

Barlas et al.[15] have developed a System Dynamics simulation model of a typical retail supply chain. The intent of their simulation exercise was to build up inventory policies that enhance the retailer's revenue and reduce costs at the same instant. Besides, the research was also intended towards studying the implications of different diversification strategies. A supply chain model functioning under periodic review base-stock inventory system to assist the manufacturing managers at HP to administer material in their supply chains has been introduced by Lee et al.[16]. The inventory levels across supply chain members were obtained with the aid of a search routine.

The inventory and supply chain managers are mainly concerned holding of the excess stock levels and hence the increase in the holding cost. Meanwhile, there is possibility for the shortage of products. For the shortage of each product there will be a shortage cost. Holding excess stock levels as well as the occurrence of shortage for products lead to the increase in the supply chain cost. The factory may manufacture any number of products, each supply chain member may consume a few or all the products and each product is manufactured using a number of raw materials sourced from many suppliers. All these factors pose additional holding of the excess stock levels and hence the increase in the holding cost. Meanwhile, there is possibility for the shortage of products. For the shortage of each product there will be a shortage cost. Holding excess stock levels as well as the occurrence of shortage for products lead to the increase in the supply chain cost. All these factors pose additional challenge in extracting the exact product and the stock levels that influence the supply chain cost heavily.

Many well-known algorithmic advances in optimization have been made, but it turns out that most have not had the expected impact on the decisions for designing and optimizing supply chain related problems. Some optimization techniques are of little use because they are not well suited to solve complex real logistics problems in the short time needed to make decisions and also some techniques are highly problem-dependent which need high expertise. This adds difficulties in the implementations of the decision support systems which contradicts the tendency to fast implementation in a rapidly changing world. IO techniques need to determine a globally optimal placement of inventory, considering its cost at each stage in the supply chain and all the service level targets and replenishment lead times that constraint each inventory location.

III. OBJECTIVES

The supply chain cost increases because of the influence of lead times for supplying the stocks as well as the raw materials. Practically, the lead times will not be same throughout all the periods. Maintaining abundant stocks in order to avoid the impact of high lead time increases the holding cost. Similarly, maintaining fewer stocks because of ballpark lead time may lead to shortage of stocks. This also

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

happens in the case of lead time involved in supplying raw materials. A better optimization methodology would consider all these above mentioned factors in the prediction of the optimal stock levels to be maintained such that the total supply chain cost can be minimized. Here, an optimization methodology that utilizes the Particle Swarm Optimization (PSO) algorithm, one of the best optimization algorithms, is proposed to overcome the impasse in maintaining the optimal stock levels in each member of the supply chain. Taking into account the stock levels thus obtained from the proposed methodology, an appropriate stock levels to be maintained in the approaching periods that will minimize the supply chain inventory cost can be arrived at.

Supply chain model is broadly divided into four stages in which the optimization is going to be performed. The supply chain model is illustrated in the Fig. 1.

A. Inventory Optimization of Multiproduct, Multiple Suppliers with Lead Time

Effective supply chain strategies must take into account the interactions at various levels in the supply chain. It is challenging to design and operate a supply chain so that total system wide costs are minimized and system wide service levels are maintained.

Uncertainty is inherent in every supply chain. Supply chains need to be designed to eliminate as much uncertainty as possible to deal effectively with the uncertainty that remains. Demand is not the only source of uncertainty. Delivery lead times, manufacturing yields, transportation times and component/raw material availability can also have significant impact on supply chain. Inventory and back order levels fluctuate considerably across the supply chain, even when customer demand for specific products does not vary greatly. The two desired objectives of improved service and inventory levels seem to be not achieved at the same time since traditional inventory theory tells us that to increase service level, the firm must increase inventory and therefore cost. But recent developments in information and communications technologies, have led to the innovative approaches that allow the firm to improve both the objectives simultaneously.

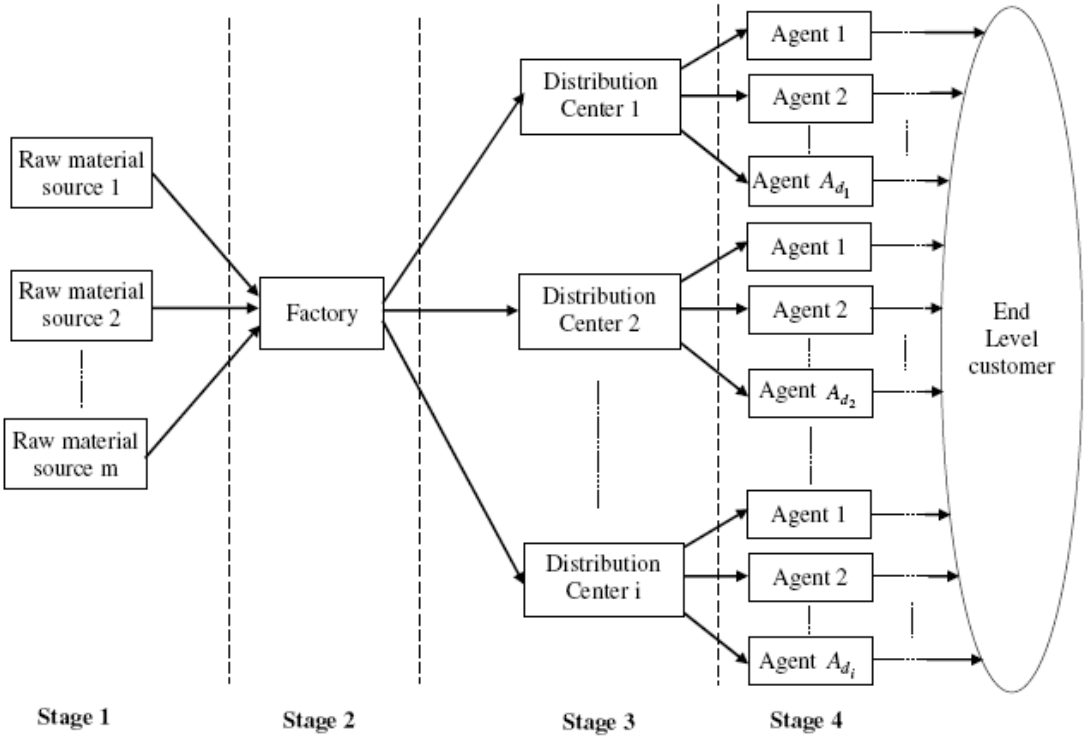


Fig. 1 Four Stage Supply Chain Model

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

In this present research, a prediction analysis that considers all these factors that are mentioned above, such that the analysis paves the way for minimizing the supply chain cost has been proposed. The supply chain cost can be minimized by maintaining optimal stock levels in each supply chain member. Such optimal stock levels can be predicted only by analyzing the past records. This minimization of supply chain cost is effective only if the optimal level is predicted with the knowledge of the lead times of the stocks. Hence a methodology is developed that analyze the past records and predict the emerging excess/shortage of stock levels that are to be considered to identify the optimal stock levels which will be maintained by each of the supply chain member.

Particle Swarm Optimization algorithms, one of the optimization algorithms in Evolutionary computation is used for analysis purpose. The stock levels that are obtained from the analysis are the stock levels that contribute more to the increase of total supply chain cost. These stock levels are used for the prediction of the optimal stock levels that need to be maintained in each supply chain member.

B. PSO Model for Prediction Analysis

The methodology proposed here will minimize the total supply chain cost by predicting optimal stock levels not only by considering the past records with respect to the stock levels, but also the lead time of the products to reach each supply chain member from its previous stage as well as the lead time taken in supplying the raw materials to the factory. Usually, shortage for a particular stock at a particular member, excess stock levels at a particular member, time required to transport stock from one supply chain member to another i.e. lead time of a stock at a member, time taken to supply raw materials to the factory to manufacture certain products i.e. lead time of raw materials in factory are some of the key factors that play vital role in deciding the supply chain cost. A better optimization methodology should consider all these factors. In the proposed methodology all the above mentioned key factors in predicting the optimal stock levels are considered. Also, different priorities are assigned to those above factors. As per the priority given, the corresponding factors will influence the prediction of optimal stock levels. Hence as per the desired requirement, the optimal stock level will be maintained by setting or changing the priority levels in the optimization procedure.

The optimization is going to be performed in the supply chain model as illustrated in the Fig. 1.

The members participating in the supply chain model are raw material sources $\{r_1, r_2, r_3, \cdots, r_m\}$, a factory f, i distribution centers $D = \{d_1, d_2, d_3, \cdots, d_i\}$ and the agents $A = \{A_{d_1}, A_{d_2}, A_{d_3}, \cdots, A_{d_i}\}$, A_{d_i} is the number of agents for the distribution center d_i. Hence, the total number of agents in the supply chain model can be arrived using formula :

$$N_A = \sum_{m=1}^{i} A_{d_m} \qquad (1)$$

where N_A is the total number of agents used in the supply chain model.

The factory is manufacturing k number of products. The database holds the information about the stock levels of each product in each of the supply chain member, lead time of products at each supply chain member and lead time of raw material. For l members from factory to end-level-Agents, there are $l-1$ lead times for a particular product and these times are collected from the past records. Similarly, the lead time for raw materials from r_m to f is also taken from the earlier period and thus the database is constituted. Each and every dataset recorded in the database is indexed by a Transportation Identification (TID). For p periods, the TID will be $\{T_1, T_2, T_3, \cdots, T_p\}$. This TID will be used as an index in mining the lead time information.

Now, the particle Swarm Optimization (PSO) is utilized to predict the emerging excess/shortage of stock levels which are vital information for optimal stock levels to be maintained in the future to minimize the supply chain cost. The procedures involved in determining the optimal stock levels are illustrated in Fig. 2.

As the particle swarm optimization (PSO) is more suitable for finding the solution for the optimization problem, PSO is utilized in finding the optimal stock levels to be maintained in each member of the supply chain. The flow of procedures is discussed below.

The individuals of the population including searching points, velocities, p_{best} and g_{best} are initialized randomly but within the lower and upper bounds of the stock levels for all supply chain members, which have to be specified earlier. Hence the generated searching point individual is

$$I_i = [P_k \quad S_1 \quad S_2 \quad S_3 \cdots\cdots S_l], \; i = 1,2,3\cdots\cdots, N_p \qquad (2)$$

where, $P_{L.B} < P_k < P_{U.B}$, $S_{L.B} < S_l < S_{U.B}$

$P_{L.B}$, $P_{U.B}$ and $S_{L.B}$, $S_{U.B}$ are the lower and upper bound values of the number of products and stock levels respectively.

The generated population is having the size of N_p i.e. the number of individuals. Since the total number of members that are maintaining the stock levels from f is l, the dimension d of each individual is given by

$$d = l + 1 \qquad (3)$$

and hence the equation (2).

Similarly, the initial velocity for the individual will be

$$[v_1 \quad v_2 \quad v_3 \cdots\cdots v_{l+1}], \; v_{min} < v_{l+1} < v_{max} \qquad (4)$$

where v_{min} and v_{max} are the minimum and maximum limit for velocities respectively.

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

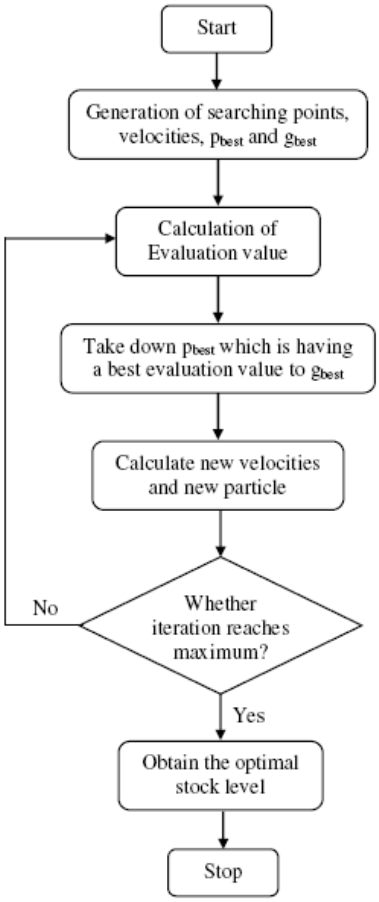


Fig. 2 Particle Swarm Optimization Steps in Optimizing the Stock Levels

Then each individual is queried into the database for obtaining the details regarding the TID and frequency of the individual. This will bring T_q, $q \in p$ and $P(occ)$, number of periods of occurrence of that particular individual. Here, q need not to be consecutive integers. This obtained TID is queried into the database having the lead time of a particular product to a particular supply chain member. The product I.D. obtained from the individual is queried into the database having the lead time each raw material for the particular product. After all these queries, the lead time of stocks obtained is as follows

$$T_s = [t_{q,1} \quad t_{q,2} \quad \cdots \quad t_{q,l-1}] \tag{5}$$

And the lead time for raw materials is obtained as

$$T_r = [t_1 \quad t_2 \cdots\cdots \quad t_r] \tag{6}$$

where r is the number of raw materials required for a certain product.

Then for each individual the evaluation function is calculated.

Determination of Evaluation function

The evaluation function is determined for each randomly generated individual. The function is given by

$$f(a) = w_1\left(1 - \frac{P(occ)}{T(periods)}\right) + \log(w_2.t_{stock} + w_3.t_{raw}) \qquad a = 1,2,3\cdots\cdots, N_p \tag{7}$$

where $T(periods)$ is the total number of periods of records in database.

In equation (7), w_1, w_2 and w_3 are the weightings of the factors, stock levels, lead time of stocks and lead time of raw materials in optimization, respectively and they are determined as

$$w_1 = \frac{R_1}{R_1 + R_2 + R_3} \tag{8.1}$$

$$w_2 = \frac{R_2}{R_1 + R_2 + R_3} \tag{8.2}$$

$$w_3 = \frac{R_3}{R_1 + R_2 + R_3} \tag{8.3}$$

R_1, R_2 and R_3 are the priority levels of influence of stock levels, lead time of stocks and lead time of raw materials in optimization respectively. Increasing the priority level of a factor increases the influence of the corresponding factor in the evaluation function. Hence this R_1, R_2 and R_3 decides the amount of influence of the factors. The lead time of the stocks t_{stock} is determined as follows

$$t_{stock} = \sum_{i=1}^{l-1} \sum_{q} t_{q,i} \tag{9}$$

And the lead time required to fill the raw materials is given as

$$t_{raw} = \sum_{i=1}^{r} t_i \tag{10}$$

Equation (9) and (10) is substituted in the equation (7) gives an evaluation value for each individual.

For every individual, a comparison is made between its evaluation value and its p_{best} .The g_{best} indicates the most excellent evaluation value among the p_{best} . This g_{best} is nothing but an index that points the best individual we have generated so far.

Subsequently the adjustment of the velocity of each particle a is as follows:

$$v_{new}(a,b) = w * v_{cnt}(a) + c_1 * r_1 * [p_{best}(a,b) - I_{cnt}(a,b)] + c_2 * r_2 * [g_{best}(b) - I_{cnt}(a,b)] \tag{11}$$

where,

$$a = 1,2,\cdots\cdots, N_p$$
$$b = 1,2,\cdots\cdots, d$$

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

Here $v_{cnt}(a)$ represents current velocity of the particle, $v_{new}(a,b)$ represents new velocity of a particular parameter of a particle, r_1 and r_2 are arbitrary numbers in the interval $[0,1]$, c_1 and c_2 are acceleration constants (often chosen as 2.0), w is the inertia weight that is given as

$$w = w_{max} - \frac{w_{max} - w_{min}}{iter_{max}} \times iter \qquad (12)$$

where w_{max} and w_{min} are the maximum and minimum inertia weight factors respectively that are chosen randomly in the interval $[0,1]$

$iter_{max}$ is the maximum number of iterations

$iter$ is the current number of iteration

Such newly obtained particle should not exceed the limits. This would be checked and corrected before proceeding further as follows,

If $v_{new}(a,b) > v_{max}(b)$, then $v_{new}(a,b) = v_{max}(b)$

if $v_{new}(a,b) < v_{min}(b)$, then $v_{new}(a,b) = v_{min}(b)$

Then as per the newly obtained velocity, the parameters of each particle is changed as follows

$$I_{new}(a,b) = I_{cnt}(a,b) + v_{new}(a,b) \qquad (13)$$

Then the parameter of each particle is also verified whether it is beyond the lower bound and upper bound limits. If the parameter is lower than the corresponding lower bound limit then replace the new parameter by the lower bound value. If the parameter is higher than the corresponding upper bound value, then replace the new parameter by the upper bound value. For instance,

If $P_k < P_{L.B}$, then $P_k = P_{L.B.}$

Similarly, if $P_k > P_{U.B}$, then $P_k = P_{U.B.}$

This is to be done for the other parameters also.

This process will be repeated again and again until the maximum number of iterations is reached. Once the maximum number of iterations is attained the process is terminated. The latest g_{best} pointing the individual is the best individual which is having the stock levels that are to be considered and these stock levels are utilized in taking the necessary steps for maintaining the optimal stock levels at each of the supply chain members.

IV. EXPERIMENTAL RESULTS

The approach suggested for the optimization of inventory level and thereby efficient supply chain management has been implemented in the MATLAB 7.4. The database consists of the records of stock levels held by each member of the supply chain for every period. For implementation purpose, five different products are utilized and these products are in circulation in the seven member supply chain network considered. The sample database which consists of the past records is shown in Table 1.

In the database tabulated in Table 1, the second field comprises of the product Identification (PI) and the other fields are related with the stock levels that were held by the respective seven members of the supply chain network. For example, the second attribute and second field of the database is '3' which refers the Product I.D. '3'. The corresponding fields of the same attribute denote the stock levels of the product I.D. '3' in the respective members of the supply chain. Similarly, different sets of stock levels are held by the database.

Table 1: Sample data from database of different stock levels

TI	PI	F1	F2	F3	F4	F5	F6	F7
1	3	632	424	247	-298	-115	365	961
2	5	-415	488	-912	979	-492	-922	205
3	2	369	-686	-468	-807	183	-386	-228
4	2	459	289	-522	-316	130	-854	468
5	3	-663	944	856	451	-763	657	484
6	4	-768	-937	-768	242	369	-890	289
7	3	-890	619	-629	-844	791	285	596
8	3	193	-263	-474	325	-409	-216	-738
9	2	578	-890	675	-321	-411	239	-916
10	1	-192	-421	593	394	-141	955	-456
11	5	494	-317	600	363	698	927	621
12	3	-838	-355	776	682	-813	-350	-513
13	5	673	-593	628	637	643	622	-244
14	1	-974	-311	-319	-189	-449	663	520
15	3	-725	797	184	406	-888	-575	-144
16	4	811	-569	-473	615	467	-748	192
17	1	-102	-703	-859	983	206	-803	-445
18	2	426	689	-735	-465	680	-913	147
19	3	655	626	-158	-485	-622	-928	515
20	5	-175	662	-847	819	239	-902	372

Table 2: Sample data from database which is having lead times for stocks

TI	T1	T2	T3	T4	T5	T6
1	28	27	19	9	19	19
2	35	33	16	4	24	15
3	38	38	20	8	10	18
4	25	25	9	21	22	13
5	45	40	15	4	16	11
6	36	43	7	13	21	3
7	31	27	20	5	22	4
8	45	28	5	8	10	6
9	41	45	6	22	8	20
10	34	26	5	21	12	4
11	45	38	1	2	11	16
12	40	38	9	18	22	23
13	38	40	11	17	7	10
14	47	47	22	8	23	23
15	35	27	16	12	22	4

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

Sample data from database which is having lead times for stocks is given in Table 2; and sample data set taken from a database which is having raw material lead time for different products is given in Table 3. Initial random individuals for running the PSO algorithm is given in Table 4; and initial random velocities corresponding to each particle of the individual required for the PSO algorithm is given in Table 5.

Table 3: Raw material lead time for different products

PI	RM	T
1	1	20
1	2	3
1	3	8
2	1	10
2	2	3
2	3	9
2	4	23
2	5	7
3	1	24
3	2	23
3	3	20
3	4	22
4	1	3
4	2	22
4	3	19
4	4	18
5	1	23
5	2	16
5	3	23
5	4	21

Table 4: Initial Random Individuals

PI	F1	F2	F3	F4	F5	F6	F7
3	855	61	215	863	24	75	-757
5	854	-154	145	-241	-215	415	845

Table 5: Initial Random Velocities corresponding to each particle of the individual

PI	F1	F2	F3	F4	F5	F6	F7
-0.1298	0.0376	-0.3439	0.3567	0.0982	-0.0560	-0.1765	-0.0409
-0.4997	0.0863	0.3573	-0.0113	0.0524	0.2177	0.6550	0.0342

V. Discussion Of Results

An iteration involving all these processes was carried out so as to obtain the best individual. Here for instance, the iteration value of '100' is chosen and so hundred numbers of iterative steps will be performed. The best individual obtained as a result is

3	-602	-280	-821	398	382	-764	-125

and its database format is depicted in the table 6.

Table 6: Database format of Final Best Individual

PI	F1	F2	F3	F4	F5	F6	F7
3	-602	-280	-821	398	382	-764	-125

As long as minimization of the fitness function is still possible, then the iteration continues till such a time that no improvement in the evaluation function value is noticeable. After a certain number of iterations, if the evaluation function value is not improving from the previous iterations, then this is an indication that the evaluation function value is stabilizing and the algorithm has converged towards optimal solution. This inference is useful for deciding the number of iterations for running the PSO simulation as well as this may be used as the stopping criteria for the algorithm.

For greater accuracy, the number of iterations should be sufficiently increased and run on the most frequently updated large database of past records. In our experimentation, the iteration value of 100 is chosen and its value is 3.8220 and the values for weighting factors w1, w2 and w3 obtained are 0.6250; 0.3125 and 0.0625 respectively.

The final individual obtained from the PSO based analysis shown in the Table 6 is the inventory level that has potential to cause maximum increase of supply chain cost. It is inferred that controlling this resultant individual is sufficient to reduce the loss either due to the holding of excess stocks or due to the shortage of stocks. By focusing on the excess/shortage inventory levels and initiating appropriate steps to eliminate the same at each member of the chain, the organization can optimize the inventory levels in the upcoming period and thus minimize the supply chain cost.

That is, the organization should focus on the most potential product among multi products that could cause maximum inventory cost and in this specific case it is product 3. The organization should take necessary steps to increase the production of product 3 in the factory by 602 units to make up for the predicted shortage; increase the inventory level of product 3 by 280 units in distribution centre 1 to make up for the predicted shortage; increase the inventory level of product 3 by 821 units in distribution centre 2 to make up for the predicted shortage; decrease inventory level of product 3 by 398 units in agent 1 to make up for the predicted excess; decrease inventory level of product 3 by 382 units in agent 2 to make up for the predicted excess; increase the inventory level of product 3 by 764 units in agent 3 to make up for the predicted shortage; increase the inventory level of product 3 by 125 units in agent 4 to make up for the predicted shortage.

Thus by following the predicted stock levels, the increase in the supply chain cost due to excess/shortage inventory levels can be avoided. The analysis provided an inventory level that had maximum potential to cause a remarkable contribution towards the increase of supply chain cost. The organization can predict the future optimal inventory levels in all the supply chain members with the aid of these levels. Therefore it is concluded that it is possible to minimize the supply chain cost by maintaining the optimal stock levels in the upcoming period among various partners in the supply chain that was predicted from the inventory analysis, making the inventory management further effective and efficient.

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

VI. CONCLUSION

Inventory management is an important component of supply chain management. As the lead time plays vital role in the increase of supply chain cost, the complexity of predicting the optimal stock levels increases. The novel and proficient approach based on PSO algorithm, one of the best optimization algorithms, is proposed to overcome the impasse in maintaining the optimal stock levels in each member of the supply chain. The proposed methodology reduced the total supply chain cost as it undoubtedly established the most probable surplus stock level and shortage level along with the consideration of lead time in supplying the stocks as well as raw materials that are required for inventory optimization.

The organizations can make use of the proposed techniques in this present research for inventory optimization by capturing the database in the desired format to suit their respective supply chain structure, replacing the simulated data used in this research with the real data of the organization. For greater accuracy, the number of iterations should be sufficiently increased and run on the most frequently updated large database of past records. Also the organization can mention the maximum possible lower limit and upper limit for the shortage and excess inventory levels respectively, within which the inventory is expected to fluctuate among the various members of the supply chain to make the convergence faster towards optimal solution. Further if the organization can evaluate and quantify the cost involved for each shortage as well as excess of inventory at each member of the supply chain, then the exact savings due to inventory optimization in the supply chain can be calculated for the organization.

VII. FUTURE SCOPE

Organization can adopt decomposition technique and use the proposed techniques for inventory optimization of high/medium/low value independent products. In this case, the organization should apply the technique for the high value items separately, medium value products separately and low value products separately for inventory optimization in the supply chain.

REFERENCES

[1]. H. Lu, 2003, "Dynamic Population Strategy Assisted Particle Swarm Optimization in Multiobjective Evolutionary Algorithm design", IEEE Neural Network Society, IEEE NNS Student Research Grants 2002, Final reports.

[2]. Hirotaka Yoshida Kenichi Kawata and Youshikazu Fukuyama Yosuke Nakanishi, 1999, "A Particle Swarm Optimization for Reactive Power and voltage control considering voltage stability", *Proceedings of IEEE International Conference on Intelligent System Applications to Power Systems*, pp. 117- 121, April 4- 8.

[3]. Mahamed G.H. Omran, Andries P Engelbrecht, and Ayed Salman, 2005, "Dynamic Clustering using Particle Swarm Optimization with Application in Unsupervised Image Classification", *Proceedings of World Academy of Science, Engineering and Technology*, Vol. 9, pp. 199-204

[4]. S. A. Hamdan, 2008, "Hybrid Particle Swarm Optimiser using multi-neighborhood topologies", *INFOCOMP - Journal of Computer Science*, Vol.7, No.1, pp.36-43.

[5]. Ling-Feng Hsieh, Chao-Jung Huang and Chien-Lin Huang, 2007, "Applying Particle Swarm Optimization to Schedule Order Picking Routes in a Distribution Center", *Asian Journal on Management and Humanity Sciences*, Vol. 1, No. 4, pp. 558- 576.

[6]. Alberto Moraglio, Cecilia Di Chio, Julian Togelius and Riccardo Poli, 2008, "Geometric Particle Swarm Optimization", *Journal on Artificial Evolution and Applications*, Article ID 143624.

[7]. T.F. Abdelmaguid and M.M. Dessouky, 2006, "A genetic algorithm approach to the integrated inventory-distribution problem", *International Journal of Production Research.*, 44: 4445-4464.

[8]. P. Pongcharoen, A. Khadwilard and A. Klakankhai, 2007, "Multi-matrix real-coded Genetic Algorithm for minimizing total costs in logistics chain network", *Proceedings of World Academy of Science, Engineering and Technology*, Vol. 26, pp. 458-463, 14-16, December.

[9]. S. Buffett and N. Scott, 2004, "An Algorithm for Procurement in Supply Chain Management", *AAMAS-04 Workshop on Trading Agent Design and Analysis*, New York.

[10]. D. Pardoe and P. Stone, 2007, An Autonomous Agent for Supply Chain Management. In: Handbooks in Information Systems Series: Business Computing, Adomavicius, G. and A. Gupta (Eds.). Elsevier. Amsterdam. http://www.cs.utexas.edu/~pstone/Papers/bib2html/b2hd-TacTex-Book07.html

[11]. B.M. Beamon, 1998, "Supply chain design and analysis: Models and methods", International Journal on Production Economics, 55: 281-294.

[12]. J.L. Caldeira, R.C. Azevedo, C.A. Silva and J.M.C. Sousa, 2007, "Supply-chain management using ACO and beam-ACO algorithms", Proceedings of the IEEE International Fuzzy Systems Conference, July 23-26, London, pp. 1-6.

[13]. K. Wang and Y. Wang, 2008, "Applying Genetic Algorithms to Optimize the Cost of Multiple Sourcing Supply Chain Systems - An Industry Case Study", Studies on Computational Intelligence, Vol. 92, pp. 355-372.

[14]. Chih-Yao Lo, 2008, "Advance of Dynamic Production-Inventory Strategy for Multiple Policies using Genetic Algorithm", *Information Technology Journal*, Vol. 7, pp. 647-653.

[15]. Y. Barlas and A. Aksogan, 1996, "Product Diversification and Quick Response Order Strategies in Supply Chain Management", 14th International Conference of the System Dynamics Society 1996 Cambridge, Massachusetts, USA, pp. 51-56.

[16]. H.L. Lee and C. Billington, 1995. The evolution of supply-chain-management models and practice at Hewlett-Packard. Interface, 25: 42-63.

[17]. K.E. Parsopoulos and M.N. Vrahatis, 2005 "Recent approaches to global Optimization problems through Particle Swarm Optimization", *Natural Computing*, 1 : 235-306.

[18]. http://swarmintelligence.org

ABOUT AUTHORS

Mrs. S. Narmadha is working as Assistant Professor in the Department of Computer Science and Engineering in Park College of Engineering and Technology, Coimbatore. She obtained her Bachelor's degree in Computer Science and Engineering from Tamilnadu College of Engineering, Coimbatore under Bharathiar University and Master's degree in Mechatronics from Vellore Institute of Technology, Vellore. She is currently pursuing Ph.D. under Anna University, Chennai. She has 8 years of Teaching Experience and 2 years of Industrial experience. She has published 14 papers in International Conferences, 2 papers in International journals and a book on 'Open Source Systems'. She is life member of ISTE. Her field of interest includes Supply Chain Management, Automation, Database Management Systems, Virtual Instrumentation, Soft Computing Techniques and Image Processing.

Dr. V. Selladurai is the Professor and Head, Department of Mechanical Engineering and Principal, Coimbatore Institute of Technology, Coimbatore, India. He holds a Bachelor's degree in Production Engineering, a Master's degree in Industrial Engineering specialisation and a Ph.D. degree in Mechanical Engineering. He has two years of industrial experience and 22 years of teaching experience. He has published over 90 papers in the proceedings of the leading National and International Conferences. He has published over 35 papers in international journals and 22 papers in national journals. His areas of interest include Operation Research, Artificial Intelligence, Optimization Techniques, Non-Traditional Optimization Techniques, Production Engineering, Industrial and Manufacturing Systems, Industrial Dynamics, System Simulation, CAD/CAM, FMS, CIM, Quality Engineering and Team Engineering.

Mr. G. Sathish is a full time research scholar under Anna University, Coimbatore. He holds a Bachelor's degree and a Master's degree in Computer Science and Engineering. He has 8 years of industrial experience. He has published 10 papers in the proceedings of the leading International Conferences. His field of interest includes Supply Chain Management, Optimization Techniques, Data Mining, Knowledge Discovery, Automation, Soft Computing Techniques and Image Processing.

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

Test Case Generation using Mutation Operators and Fault Classification

Mrs. R. Jeevarathinam [#1], Dr. Antony Selvadoss Thanamani [*2]

[#] *Department of Computer Science*
SNR Sons College, Coimbatore, Tamilnadu, India.
[1]

[*] *Associate Professor and Head*
Department of Computer Science
NGM College, Pollachi, Tamilnadu, India.
[2]

***Abstract*— Software testing is the important phase of software development process. But, this phase can be easily missed by software developers because of their limited time to complete the project. Since, software developers finish their software nearer to the delivery time; they don't get enough time to test their program by creating effective test cases. . One of the major difficulties in software testing is the generation of test cases that satisfy the given adequacy criterion Moreover, creating manual test cases is a tedious work for software developers in the final rush hours. A new approach which generates test cases can help the software developers to create test cases from software specifications in early stage of software development (before coding) and as well as from program execution traces from after software development (after coding). Heuristic techniques can be applied for creating quality test cases. Mutation testing is a technique for testing software units that has great potential for improving the quality of testing, and to assure the high reliability of software. In this paper, a mutation testing based test cases generation technique has been proposed to generate test cases from program execution trace, so that the test cases can be generated after coding. The paper details about the mutation testing implementation to generate test cases. The proposed algorithm has been demonstrated for an example.**

I. INTRODUCTION

Software testing is a standard method of assuring software quality. Software testing is an important activity to assure the quality of software. Unfortunately, software testing is very labor intensive and very expensive. It can take about 50 percents of total cost in software developing process [1]. The software testers may need to spend a longer time using many test cases if the test data used are not of high quality. Therefore, a performance of executing test case is an important issue to reduce the testing time. Software testing is usually the first part of software development stages, which software developers decide to omit when there is a limited time to deliver the software. In other word, developers may not have enough time after they finished their coding to create test cases to test their code. Generating test cases can resolve these problems. This not only helps developers to test their program when they finish coding but also controls the developers to program the software as defined in the software specification [2]. In this case the software specifications are the main sources for generating test cases as these documents describe the software system to be developed in detail.

One of the most difficult and expensive parts of applying these techniques has been the actual generation of test data- which has traditionally been done by hand

The general aim of the research reflected in this paper is to formalize, and mechanize where possible, routine aspects of testing. Such formalization has two benefits. First, it makes it easier to analyse a given test set to ensure that it satisfies a specified coverage criterion. Second, it frees the test engineer to concentrate on less formalization, and often more interesting tests. Developers have responded to this need in many ways, including improving the process, increasing the attention on early development activities, and using formal methods to describe requirements, specifications, and designs. Although all of these improvements help create software that is of higher quality and higher reliability, the software still needs to be tested, and the more stringent needs for the product also means that the testing method must be more effective at finding problems in the software. Project and test managers are more than ever in a position where they need solid information for how to apply scarce resources. Applying structured, precisely defined testing techniques allows development resources to be used more wisely.

Specification-based testing refers to creating test inputs from the software specifications. Specification-based testing allows tests to be created earlier in the development process, and be ready for execution before the program is finished. Additionally, when the tests are generated, the test engineer will often find inconsistencies and ambiguities in the specifications, which allows problems to be found and eliminated early. Specifications can be used as a basis for output checking, which significantly reduces one of the major costs of testing. Another advantage is that the essential part of the test data can be independent of any particular implementation of the specifications. Specification-based testing is also important for conformance testing, where access to the code is not provided, but specifications for the product are.

There is an increasing need for effective testing of software for growing applications, such as web applications

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

and e-commerce require software that exhibits more reliability than most traditional application areas. Without software that functions reliably, businesses that operate on the web will lose money, sales, and customers. In recent years, the phrase “fault-based testing" has been applied to techniques that choose test data that attempt to show the presence (or absence) of specific faults during unit testing. Techniques have been developed that determine whether test data can detect specific faults (i.e., mutation analysis [3]), and the theoretical properties of fault-based testing have been explored [5, 4].

The mutation testing is a fault based testing strategy that measures the quality of testing by examining whether the test set, test input data used in testing can reveal certain type of faults. Mutation testing helps testers create test data by interacting with them to strengthen the quality of the test data. Faults are introduced into programs by creating many versions of the software, each containing one fault. Test cases are used to execute these faulty programs with the goal of causing each faulty program to produce incorrect output (fail). Hence the term mutation; faulty programs are mutants of the original, and a mutant is killed when it fails. When this happens, the mutant is considered dead and no longer needs to remain in the testing process because the faults represented by that mutant have been detected.

II. RELATED WORK

If testers want to test functional requirements, they may use black-box testing technique. Black-box testing [6] does not need knowledge of how software is programmed. Test oracles are specified by software design or software specifications. Testers inject test data to execute program, then compare actual result with the specified test oracle. By contrast, white-box testing needs knowledge of how software is programmed. In white-box testing, paths or statements which has been executed are test oracle. These are called coverage criteria. There are three main types of coverage criteria: statement, coverage, branch coverage, and path coverage. Statement coverage reports whether each statement is encountered by the test suite or not. Branch coverage reports whether every branch structure (if – else clause or while clause) has been executed for true and false condition in each branch. Finally, path coverage reports whether all possible paths in function has been tested.

In Object-oriented context, the structure of software is more complicated than the structural one. Conventional test approaches may not be enough for testing. The combination of those two traditional approaches is called Gray-box testing [7]. In Gray-box testing, test data generates based on the high level design which specifies the expected structure and behaviour of system. Gray-box testing investigates the coverage criteria of white-box method and finds all possible coverage paths. In addition, the generated test case should be satisfied with functional requirement as in the black-box testing criteria.

Many automated test case generation techniques produce test cases based on Gray-box method. Not only does Gray-box testing concern functional requirement as black box testing, but also concerns on behaviours of system. Clarke [8] proposed an empirical study which compared efforts between automate test generation and manual test generation. In his report test data was generated from extended finite state model (EFSM). The research shows that the automate test data generation could reduce an effort from manual test data generation for more than 88 percents. Xu and Yang [9] proposed test data generation framework called JMLAutoTest framework. JMLAutoTest framework generates test data from Java Modelling Language (JML) [[10] [11]]. JML is a notation for specifying behaviour and interface in Java class and method. Since JML is a formal specification, developers should spend efforts to understand JML before writing specification. Because UML diagrams are now widely used for software development [12], generating test data from UML diagrams should help developer to reduce a great number of efforts.

Wang, et.al [13] proposed test data generation from activity diagram. They extracted a test scenario from activity diagram. The test scenario is a sequence of possible paths in activity diagram. From these paths, the executing sequence of program has been generated in order to cover all possible paths. However, activity diagram describes flows of system, not the behaviour of the system. Due to performance of generating test data and a concern of size of test data set, heuristic techniques are applied for test data generation. GADGET [14] and TGEN [15] use genetic algorithm to improve quality of generating test data. GADGET generates test data from a control flow graph generated from source code. A fitness function is defined for each condition node in control flow graph. An empirical study showed that test data generated by GADGET covers more than 93 percents of source code, while random testing achieves around 55 percents. TGEN transforms a control flow graph to a control dependency graph (CDG). Each part of CDG represents the smallest set of predicate to traverse every node in control flow graph. Both GADGET and TGEN generate test data using white box method; therefore, test data can be generated only after software is finished. Using Genetic algorithm to generate test data from software model is proposed in [16]. JML is a model for generating test data. Fitness function is calculated by coverage of paths and post condition defined by JML.

Because of the large number of mutant programs that must be generated and run, early designers of mutation analysis systems considered individually creating, compiling, linking, and running each mutant more difficult, and slower, than using an interpretive system [[17] [18]] . It was considered likely that the cost of compiling large numbers of mutants would be prohibitive. Of the interpreter-based systems that have been developed, Mothra is the most recent and comprehensive [[19] [21]]. In these conventional, interpreter-based mutation analysis systems, the source code is translated into an internal form suitable for interpretive execution and mutation. For each mutant, a mutant generator program produces a “patch" that, when applied to the internal form, creates the desired alternate program. The translated

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

program plus the collection of patches represents a program neighbourhood. To run a mutant against a test case, the interpreter dynamically applies the appropriate patch and interpretively executes the resulting alternate internal form program. A number of attempts to overcome the performance problem have been made. Some approaches attempt to limit the number of mutants that must be run. In selective mutation [26], only a subset of the possible mutagens is used, resulting in fewer mutants being created. Preliminary results suggest that selective mutation may provide almost the same test coverage as non-selective mutation under certain conditions. Running only a sample of the mutants [27] has also been suggested.

In extreme cases, however, it is necessary to run almost all the mutants. In other approaches, the use of non-standard computer architectures has been explored. Unfortunately, full utilization of these high performance computers requires an awareness of their special requirements as well as adaptation of software. Work has been done to adapt mutation systems to vector processors [23], to SIMD [22] and hypercube (MIMD) machines [[24] [25]]. However, it is the very fact that these architectures are non-standard that limits the appeal of these approaches. Not only are they not available in most development environments, but testing software designed for one operational environment (machine, operating system, compiler, etc.) on another is fraught with risks.

The approaches above do not squarely address the primary factor that causes conventional systems to be slow: interpretative execution. As noted previously, the overhead of compiling many mutant programs outweighs the benefit of increased execution speed. Compiler-integrated [20] program mutation seeks to avoid excessive compilation overhead and yet retain the benefit of compiled speed execution. In this method, the program under test is compiled by a special compiler. As the compilation process proceeds, the effects of mutations are noted and code patches that represent these mutations are prepared. Execution of a particular mutant requires only that the appropriate code patch be applied prior to execution. Patching is inexpensive and the mutant executes at compiled-speeds. Unfortunately, crafting the needed special compiler is an expensive undertaking. Modifying an existing compiler reduces this burden somewhat, but the task is still technically demanding. Moreover, for each new computer and operating system environment, this task must be repeated.

III. FAULT CLASSIFICATION

A test case that distinguishes the program from its mutant is considered to be effective at finding faults in the program. The effectiveness of mutation testing, like other fault-based approaches, depends heavily on the types of faults that the mutation system is designed to represent. Since mutation testing uses mutation operators to implement faults, the quality of the mutation operators is crucial to the effectiveness of mutation testing. Although mutation testing has a rich history, most mutation operators have been developed for procedural programs. OO languages contain new features such as encapsulation, inheritance, and polymorphism. These features introduce the potential for new faults. Therefore, existing mutation operators for procedural programming languages are not sufficient for programs written in OO languages and new OO-specific language operators are needed.

The effectiveness of mutation testing depends heavily on the types of faults that may be represented. In these new kinds of faults, some of which are not modelled by traditional mutation operators. Which are insufficient to test these OO language features, particularly at the class testing level. This paper introduces a new set of class mutation operators for the OO languages. These operators are based on specific OO faults and can be used to detect faults involving inheritance, polymorphism, and dynamic binding, thus are useful for inter-class testing. The faults modelled by these operators are not general; they can be application-specific or programmer-specific. Therefore, to execute mutation testing with these operators, they should be selected based on the characteristic of the program to be tested. The previous attempts suffered from not having a general fault model. The previous OO mutation operators do not handle several fault types and did not handle all OO features. Faults can be classified as occurring at the intra-method level, inter-method level, intra-class level, and inter-class level.

Intra-method level faults occur when the functionality of a method is implemented incorrectly. A method in a class corresponds to the unit of the conventional program testing. Inter-method and intra-class level faults are made at the interactions between pairs of methods of a single class or between pairs of methods that are not part of a class construct in non-OO languages. Because methods are getting smaller and interactions among methods are increasingly encoding the design complexity. Inter-class level faults include faults that occur due to the object-oriented specific features such as encapsulation, inheritance, polymorphism, and dynamic binding.

IV. MUTATION OPERATORS

There are three kinds of mutation operators available namely statement level operators, method level operators and class level operators.

Statement level mutation operators involve the creation of a set of mutant programs of the program being tested. Each mutant differs from the original program by one mutation. A mutation is a single syntactic change that is made to a program statement.

Operand Replacement Operators (ORO) - Replacing a single operand with another operand or constant.

Expression Modification Operators (EMO) – Replacing an operator or inserting a new operator.

Statement Modification Operators (SMO) – Replacing or deleting a statement or part of the statement.

Method level mutation operators are used in unit and integration level testing and can be classified into two levels: (1) intra-method, (2) inter-method. This classification follows

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

definitions by Harrold and Rothermel [25] [27] and Gallagher and Offutt [26] [23].

Intra-method level faults occur when the functionality of a method is implemented incorrectly. Testing within classes corresponds to unit testing in conventional programs. So far, researchers have assumed that traditional mutation operators for procedural programs will suffice for this level (with minor modifications to adapt to new languages).

Inter-method level faults are made on the connections between pairs of methods of a single class. Testing at this level is equivalent to integration testing of procedures in procedural language programs. Interface mutation which evaluates how well the interactions between various units have been tested, is applicable to this level.

Class level mutation operators can be classified into two levels: (1) intra-method, (2) inter-method.

Intra-class testing is when tests are constructed for a single class, with the purpose of testing the class as a whole. Intra-class testing is a specialization of the traditional unit and module testing. It tests the interactions of public methods of the class when they are called in various sequences. Tests are usually sequences of calls to methods within the class, and include thorough tests of public interfaces to the class.

Inter-class testing is when more than one class is tested in combination to look for faults in how they are integrated. Inter-class testing specializes the traditional integration testing and seldom used subsystem testing, where most faults related to polymorphism, inheritance, and access are found.

Based on the fault classification, Ma et al. [28] developed a comprehensive set of class mutation operators for Java. There are 24 mutation operators explained below. Each mutation operator is related to one of the following six language feature groups. The first four groups are based on language features that are common to all object oriented languages. The fifth group includes language features that are Java-specific, and the last group of mutation operators are based on common object oriented programming mistakes. As is usual with mutation operators, they are only applied in situations where the mutated program will still compile.

A. Information Hiding

Access control is one of the common sources of mistakes among object oriented programmers. The semantics of the various access levels are often poorly understood, and access for variables and methods is not always considered during design. Poor access definitions do not always cause faults initially, but can lead to faulty behaviour when the class is integrated with other classes, modified, or inherited from. The Access Control mutation operator, Access modifier change (AMC) has been developed for this category.

B. Inheritance

Although inheritance is a powerful and useful abstraction mechanism, incorrect use can lead to a number of faults. Seven mutation operators have been defined to test the various aspects of using inheritance, covering variable hiding, method overriding, the use of super, and definition of constructors and are listed below.

IHD-Hiding variable deletion IHI-Hiding variable insertion
IOD-Overriding method deletion IOP- Overriding method calling position change
IOR-Overriding methods rename ISK-Super keyword deletion
IPC-Explicit call of a parent's constructor deletion

C. Polymorphism

Polymorphism and dynamic binding allow object references to take on different types in different executions and at different times in the same execution. That is, object references may refer to objects whose actual types differ from their declared types. In most languages (including Java and C++), the actual type can be any type that is a subclass of the declared type. Polymorphism allows the behaviour of an object reference to differ depending on the actual type. Four operators have been developed for this category.

PNC- new method call with child class type
PMD- Instance variable declaration with parent class type
PPD -Parameter variable declaration with child class type
PRV- Reference assignment with other comparable type

D. Overloading

Method overloading allows two or more methods of the same class or type family to have the same name as long as they have different argument signatures. Just as with method overriding (polymorphism), it is important for testers to ensure that a method invocation invokes the correct method with appropriate parameters. Four mutation operators have been defined to test various aspects of method overloading.

OMR- Overloading method contents change
OMD- Overloading method deletion
OAO- Argument order change
OAN- Argument number change

E. Java Specific Features

Because mutation testing is language dependent, mutation operators need to reflect language-specific features. Java has a few object-oriented language features that do not occur in all object oriented languages and four operators have been defined to ensure correct use of these features. They cover use of this, static, default constructors and initialization.

JTD- this keyword deletion
JSC- static modifier change
JID- Member variable initialization deletion
JDC-Java-supported default constructor creation

F. Common Programming Mistakes

This category attempts to capture typical mistakes that programmers make when writing object oriented software. These are related to use of references and using methods to access instance variables. Four operators have been developed for this category.

EOA- Reference assignment and content assignment replacement

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

EOC- Reference comparison and content comparison replacement
EAM- Accessor method change
EMM- Modifier method change

V. RELATIONSHIP BETWEEN FAULTS AND OPERATORS

The following table relates the fault types and our mutation operators. All faults are covered, and some required multiple mutation operators. Conversely, some of the mutation operators cover more than one fault

Faults	**Class Mutation Operators**
State visibility anomaly	IOP
State definition inconsistency (due to state variable hiding)	IHD, IHI
State definition anomaly (due to overriding)	IOD
Indirect inconsistent state definition	IOD
Anomalous construction behaviour	IOR, IPC, PNC
Incomplete construction	JID, JDC
Inconsistent type use	PID, PNC, PPD, PRV
Overloading methods misuse	OMD, OAO, OAN
Access modifier misuse	AMC
Static modifier misuse	JSC
Incorrect overloading-methods implementation	OMR
Super keyword misuse	ISK
This keyword misuse	JTD
Faults from common programming mistakes	EOA, EOC, EAM, EMM

VI. EFFECTIVENESS OF MUTATION OPERATORS

TCAS/Siemens has an internal state which is large relative to the number of inputs and outputs. TCAS, aircraft collision avoidance, is a part of a set of C programs that came originally from Siemens Corporate Research and was subsequently modified by Rothermel and Harrold [26]. These programs are used in research on program testing, so they come with extensive test suites and sets of faulty versions. There are 12 input variables specifying parameters of own aircraft and another aircraft and one output variable, alt_sep, a resolution advisory to maintain safe altitude separation between the two aircrafts. The program computes intermediate values and prints alt_sep to the standard output. The program has minimal documentation, and we wrote a formal specification for it. The following table gives the results of various mutation operators in terms of number of mutants generated, number of traces produced and the percentage of fault coverage.

There are several issues that need to be considered to evaluate the usefulness and effectiveness of the OO class level mutation operators. First is the issue of equivalent mutants. Equivalent mutants do not affect the semantics of the program; therefore they are useless for mutation testing. Second, some operators can generate mutants that are easily killed. Although mutants make simple syntactic changes to the program, their semantic impacts can vary greatly. The impacts of OO operators on the semantics can vary from affecting the method to the semantics of the entire class. For example, the IOD operator swaps overriding methods with its parent's. The effect of IHD, on the other hand, extends over the whole class because it handles instance variable, which determine the class state. It is possible that some of these mutation operators will create mutants that are too easily killed. Finally, the mutation operators need to be evaluated in terms of their effectiveness of detecting faults in OO programs. The AMC and JSC operators produced a lot of equivalent mutants, and the PNC, PMD, PPD and IHI operators produced equivalent mutants when overriding was present.

TABLE I
RESULT OF MUTATION OPERATORS AND THEIIR FAULT COVERAGE

Operator	No.of Mutants	No.of Traces	Coverage
AMC	202	24	96.6%
IOD	72	21	87.9%
ISK	130	21	93.1%
IHD	116	14	62.9%
PNC	74	18	94.2%
PPD	72	21	90.7%
PMD	144	29	83.7%
JTD	12	4	52.4%
JSC	83	17	85.2%
IHI	97	22	76.4%

VII. CONCLUSION

This paper presents a comprehensive set of mutation operators to test for faults in the use of object-oriented features. These mutation operators are based on an exhaustive list of OO faults, which gives them a firm theoretical basis. As a result, they correct several problems. These mutation operators are designed with an emphasis on the integration aspects of Java to support interclass level testing, and will help testers find faults with the use of language features such as access control, inheritance, polymorphism and overloading. Thus, this provides a way to improve the reliability of OO software.

REFERENCES

[1] Myers, G., The Art of Software Testing. 2 ed. 2004: John Wiley & Son. Inc. 234
[2] Beck, K., Test-Driven Development by Example. 2003: Addison- Wesley. 220.
[3] R. A. DeMillo, R. J. Lipton, and F. G. Sayward. Hints on test data selection: Help for the practicing programmer. IEEE Computer, 11(4):34{41, April 1978.
[4] L. J. Morell. A Theory of Error-Based Testing. PhD thesis, University of Maryland, College Park MD, 1984. Technical Report TR-1395. Rel-work-mutation testing

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

[5] T. A. Budd and D. Angluin. Two notions of correctness and their relation to testing. Acta Informatica, 18(1):31{45, November 1982.
[6] Beizer, B., Black-box testing : techniques for functional testing of software and systems. 1995: John Wiley & son Inc. 294.
[7] Hung, N.Q., Testing Application on the Web. 2003: John Wiley & Sons.
[8] Clark, J.M. Automated Test Generation from a Behavioral Model. In the 11th International Software Quality Week (QW98). 1998.
[9] Xu, G. and Z. Yang. JMLAutoTest: A Novel Automated Testing Framework Based on JML and JUnit. in Lecture Notes in Computer Science. 2004.
[10] Burdy, L., et al. An overview of JML tools and applications. in Eighth International Workshop on Formal Methods for Industrial Critical Systems (FMICS '03), ser. Electronic Notes in Theoretical Computer Science. 2003. Elsevier.
[11] Leavens, G.T., et al., JML Reference Manual. 2005.
[12] Lange, C.F.J., M.R.V. Chaudron, and J. Muskens, In practice: UML software architecture and design description. Software, IEEE, 2006. 23(2): p. 40-46.
[13] Wang, L., et al. Generating test cases from UML activity diagram based on Gray-box method. in Software Engineering Conference, 2004. 11th Asia-Pacific 2004.
[14] Michael, C., G. McGraw, and M.A. Schatz, Generating software test data by evolution. Software Engineering, IEEE Transactions on, 2001. 27(12): p. 1085-1110.
[15] Pargas, R., M. Harrold, and R. Peck, Test-data generation using genetic algorithms. Software Testing, Verification and Reliability, 1999. 9(4): p. 263-282.
[16] Cheon, Y., M.Y. Kim, and A. Perumandla. A Complete Automation of Unit Testing for Java Programs. in Proceedings of the 2005 International Conference on Software Engineering Research and Practice (SERP '05). 2005. Las Vegas, Nevada, USA,.
[17] Timothy A. Budd. Private correspondence, February 24 1992
[18] Timothy A. Budd, Richard J. Lipton, Frederick G. Sayward, and Richard A. DeMillo. The Design of a Prototype Mutation System for Program Testing. In Proceedings of the National Computer Conference, pages 623-627,.
[19] Richard A. DeMillo, Dany S. Guindi, Kim N. King, W. Michael McCracken, and A. Jefferson Offutt. An Extended Overview of the Mothra Software Testing Environment. In Proceedings of the Second Workshop on Software Testing, Veri_cation, and Analysis, pages 142-151, Ban_, Alberta, Canada, July 19{21 1988. IEEE Computer Society Press.
[20] Richard A. DeMillo, Edward W. Krauser, and Aditya P. Mathur. Compiler-Integrated Program Mutation. In Proceedings of the Fifteenth Annual International Computer Software and Applications Conference (COMPSAC), pages 351{356, Tokyo, Japan, September 11{13 1991. IEEE Computer Society Press
[21] Kim N. King and A. Jefferson Offutt. A Fortran Language System for Mutation-based Software Testing. Software-Practice and Experience, 21(7):685{718, July 1991.
[22] Edward W. Krauser, Aditya P. Mathur, and Vernon J. Rego. High Performance Software Testing on SIMD Machines. IEEE Transactions on Software Engineering, SE-17(5):403- 423, May 1991.
[23] Aditya P. Mathur and Edward W. Krauser. Mutant Unification for Improved Vectorization. Technical Report SERC-TR-14-P, Software Engineering Research Center, Purdue University, West Lafayette, IN, April 25 1988.
[24] ByoungJu Choi and Aditya P. Mathur. High Performance Mutation Testing. The Journal of Systems and Software, 20(2):135{152, February 1993.
[25] A. Jefferson Offutt, Roy P. Pargas, Scott V. Fichter, and Prashant K. Khambekar. Mutation Testing of Software Using a MIMD Computer. In Proceedings of the 1992 International Conference on Parallel Processing, pages II{257{266, St. Charles, IL, August 17{21 1992.
[26] A. Jefferson Offutt, Gregg Rothermel, and Christian Zapf. An Experimental Evaluation of Selective Mutation. In Proceedings of the Fifteenth International Conference on Software Engineering, Baltimore, MD, May 17{21 1993. IEEE Computer Society Press.
[27] Mehmet Spahinoglu and Eugene H. Spafford. A Sequential Statistical Procedure in Mutation-Based Testing. In Proceedings of the 28th Annual Spring Reliability Seminar, pages 127{148, Boston, MA, April 19 1990. Central New England Council of IEEE.
[28] Y. S. Ma, Y. R. Kwon, and J. Offutt. Inter-class mutation operators for Java. In IEEE Computer Society Press, editor, 13th International Symposium on Software Reliability Engineering, pages 352-363, Annapolis MD, November 2002.

AUTHORS PROFILE

Mrs R. Jeevarathinam graduated with MCA in 2001 from Bharathiar University, India and completed M.Phil from Bharathidasan University, India during 2003-04. Her areas of Interest include Software Engineering & Data Mining. She has about 8 years of teaching experience. Currently she is working as a Sr. Lecturer in CS department at SNR Sons College, Coimbatore, India and also pursuing PhD of Mother Teresa Women University, India. She has published a number of papers in various national & international journals & conferences. She is an active IEEE student member

Dr.Antony Selvadoss Thanamani is presently working as Reader in the Dept of Computer Science, NGM College, India. He has published more than twenty papers in national/journals and more than ten books. His areas of interest includes E-Learning, Software Engineering, Data Mining, Networking and etc. He has about 20 years of teaching experience. He is guiding many research scholars and has published many papers in national and international conference and in many international journals

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

An Energy Efficient and Reliable Congestion Control Protocol For Multicasting In Mobile Adhoc Networks

Dr.G.Sasi Bhushana Rao
Senior Professor
Department of Electronics and Communication Engineering
Andhra University
Visakhapatnam

M.RajanBabu
Associate Professor
Department of Electronics and Communication Engineering
Lendi Institute of Engineering and Technology
Jonnada, Vizianagaram, AndhraPradesh, India

Abstract— **This paper presents an energy efficient and reliable congestion control protocol for multicasting in mobile adhoc networks (MANETs). Our proposed scheme overcomes the disadvantages of existing multicast congestion control protocols which depend on individual receivers to detect congestion and adjust their receiving rates. In the first phase of our protocol, we build a multicast tree routed at the source, by including the nodes with higher residual energy towards the receivers. In the second phase, we propose an admission control scheme in which a multicast flow is admitted or rejected depending upon on the output queue size. In the third phase, we propose a scheme which adjusts the multicast traffic rate at each bottleneck of a multicast tree. Because of the on-the-spot information collection and rate control, this scheme has very limited control traffic overhead and delay. Moreover, the proposed scheme does not impose any significant changes on the queuing, scheduling or forwarding policies of existing networks. Simulation results shows that our proposed protocol has better delivery ratio and throughput with less delay and energy consumption when compared with existing protocol.**

Keywords-Congestion Control; Mobile Adhoc Networks; Multicasting; admission control; multicast tree.

I. INTRODUCTION

A mobile ad-hoc network (MANET) is composed of mobile nodes without any infrastructure. Mobile nodes self-organize to form a network over radio links. The goal of MANETs is to extend mobility into the realm of autonomous, mobile and wireless domains, where a set of nodes form the network routing infrastructure in an ad-hoc fashion. The majority of applications of MANETs are in areas where rapid deployment and dynamic reconfiguration are necessary and wired network is not available. These include military battlefields, emergency search, rescue sites, classrooms and conventions, where participants share information dynamically using their mobile devices. These applications lend themselves well to multicast operations [1].

Multicasting is aimed to deliver data to a set of selected receivers. There is no restriction on the location or number of members in a host group. Multicast can be classified into one to many or many to many communication applications. The important member identifications and functions are: group member, sources, destination, forwarding nodes, non-group member. The group membership is dynamic means that hosts may join and leave groups at any time Multicast packets are delivered to each member of a multicast group with the same best-efforts reliability and performance as unicast packets to members. Multicast groups may be of arbitrary size, may change membership dynamically, and may have either a global or local scope. The senders do not need to know membership groups, and needs not to be a member of that group. [2]. In addition, within a wireless medium, it is crucial to reduce the transmission overhead and power consumption. Multicasting can improve the efficiency of the wireless link when sending multiple copies of messages by exploiting the inherent broadcast property of wireless transmission. Hence, reliable multicast routing plays a significant role in MANETs [1].

Multicasting can be used to improve the efficiency of the wireless link when sending multiple copies of messages to exploit the inherent broadcast nature of wireless transmission. So multicast plays an important role in MANETs Unlike typical wired multicast routing protocols, multicast routing for MANETs must address a diverse range of issues due to the characteristics of MANETs, such as low bandwidth, mobility and low power. MANETs deliver lower bandwidth than wired networks; therefore, the information collection during the formation of a routing table is expensive [1].

A. Multicast Issues in MANET

Scalability: A multicast routing protocol is scalable with respect to some constraints posed by MANETs.

Multicast service support: The multicast protocol defines conditions for joining/leaving groups, multicast participants should be able to join or leave groups at will. On the other hand, service providers can be convinced to support multicast protocols.

Traffic control: Both source and core-based approaches concentrate traffic on a single node. In stateless multicast group membership is controlled by the source, which leads to the vulnerability of multicast protocols for MANETs. Still need to be investigated is how to efficiently distribute traffic from a central node to other member nodes for MANETs.

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

QoS: QoS defines a guarantee given by the network to satisfy a set of predetermined service performance constraints for the user in terms of end-to-end delay, jitter, and available bandwidth. Therefore, multicast routing protocols must be feasible for all kinds of constrained multicast applications to run well in a MANET. However, it is a significant technical challenge to define a comprehensive framework for QoS support, due to dynamic topology, distributed management and multi-hop connections for MANETs.

Multiple sources: Most of the existing multicast routing protocols in ad-hoc networks are designed for single source multicasting. However, a multicast group may contain multiple sources due to different kinds of services or applications simultaneously provided by the networks. Each single source multicast routing protocol induces a lot of overhead and thus wastes tremendous network resources in a multi-source multicast environment.

The QAMNet [9] depends on the traffic pattern hence it is difficult to accurately estimate the threshold rate. The protocols which support QoS for multicasting introduce network state and additional signaling. Such additional signaling packets for reservation protocol must be avoided as this adds to network congestion, especially in high mobility scenarios.

The QMR and E-QMR protocols calculate approximately the available bandwidth based on the channel status. This results in some problem. Each node can listen to the channel to determine the channel status and computes the idle duration only for a period of time [10]. A lantern-tree topology is used to provide QoS multicast routing. Need for a centralized MAC scheme in ad hoc mobile networks with dynamic wireless environments is its main disadvantage [12].

B. Proposed Solution

In this paper, we propose to design an energy efficient and reliable congestion control (EERCCP) protocol for multicasting with the following phases.

In its first phase, it builds a multicast tree routed at the source, by including the nodes with higher residual energy towards the receivers. Most of the existing schemes depend on individual receivers to detect congestion and adjust their receiving rates which are much disadvantageous. In the second phase, we propose an admission control scheme in which a multicast flow is admitted or rejected depending upon on the output queue size.

In the third phase, we propose a scheme which adjusts the multicast traffic rate at each bottleneck of a multicast tree.

II. RELATED WORK

Hua Chen, Baolin Sun [3] introduces an Entropy-based Fuzzy controllers QoS Routing algorithm in MANET (EFQRM). The key idea of EFQRM algorithm is to construct the new metric-entropy and fuzzy controllers with the help of entropy metric to reduce the number of route reconstruction so as to provide QoS guarantee in the ad hoc network.

Tolga Numanoglu and Wendi Heinzelman [4] propose a mesh networking inspired approach that adapts the amount of redundancy according to the current link conditions. They simultaneously reduce unnecessary energy dissipation.

D. Agrawal, T. Bheemarjuna Reddy, and C. Siva Ram Murthy [5] propose a Robust Demand-driven Video Multicast Routing (RDVMR) protocol. Their protocol uses a novel path based Steiner tree heuristic to reduce the number of forwarders in each tree. They construct multiple trees in parallel with reduced number of common nodes among them. Moreover, unlike other on-demand multicast protocols, RDVMR specifically attempts to reduce the periodic (non on-demand) control traffic.

Guojun Wang, Jiannong Cao, Lifan Zhang, Keith C. C. Chan [6] proposes a logical Hypercube-based Virtual Dynamic Backbone (HVDB) model for QoS-aware multicast communications. In this model, high fault tolerance and small diameter of hyper cubes are the basis for high availability, and regularity and symmetry of hyper cubes contribute to good load balancing.

Vida Lashkari B. O, Mehdi Dehghan [7] proposes an efficient algorithm named is proposed to improve the route discovery mechanism in MAODV for QoS multicast routes. QoS-MAODV especially can establish a multicast tree with the minimum required bandwidth support and decrease the end-to-end delay between each destination and the source node. It can establish QoS routes with the reserved bandwidth on per chosen flow. To perform accurate resource reservation, they have developed a method for estimating the consumed bandwidth in multicast trees by extending the methods proposed for unicast routing.

Zeyad M. Alfawaer, GuiWei Hua, and Noraziah Ahmed [8] introduced MANHSI (Multicast for Ad hoe Network with hybrid Swarm Intelligence) protocol, which relies on a swarm intelligence based optimization technique to learn and discover efficient multicast connectivity. The proposed protocol instances that it can quickly and efficiently establish initial multicast connectivity and/or improved the resulting connectivity via different optimization techniques.

Harald Tebbe and Andreas J. Kassler [9] present QAMNet, an approach to improve the Quality of Service (QoS) for multicast communication in MANETs. They extend existing approaches of mesh based multicasting by introducing traffic prioritization, distributed resource probing and admission control mechanisms, adaptive rate control of non-real-time traffic based on Medium Access Control (MAC) layer feedback so as to maintain low delay and required throughput for real-time multicast flows.

Mohammed Saghir, Tat-Chee Wan, Rahmat Budiarto [10] has extended QMR to make it more effective than the previous work. They propose a cross-layer framework to support QoS multicasting. They have enhanced the IEEE 802.11 MAC layer to estimate the available bandwidth at each node.

Ravindra Vaishampayan, J.J. Garcia-Luna-Aceves [11] proposed a protocol for unified multicasting through announcements (PUMA) in ad-hoc networks, which establishes and maintains a shared mesh for each multicast group, without requiring a unicast routing protocol or the pre assignment of cores to groups. PUMA achieves a high data

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

delivery ratio with very limited control overhead, which is almost constant for a wide range of network conditions.

N. Ben Ali, A. Belghith, J. Moulierac, M. Molnar [13] has proposed a new algorithm coined mQMA. This deals with two main problems of traditional IP multicast, which are multicast forwarding state scalability and multi-constrained QoS routing. The algorithm mQMA is a QoS multicast aggregation algorithm which handles multiple additive QoS constraints. It builds few trees and maintains few forwarding states for the groups. The multicast tree aggregation technique, allows several groups to share the same delivery tree. The mQMA algorithm builds trees satisfying multiple additive QoS constraints.

III. Energy Efficient And Reliable Congestion Control Protocol

A. Energy Efficient Tree Construction

In our energy efficient and reliable congestion control protocol we build a multicast tree routed at the source towards the receivers. The distance i.e. the geographical location of the nodes is assumed. Their residual energy is measured. The nodes are sorted based on its location from the source and arranged in a sequence order. A threshold value Q is set and the nodes which are less than $Q(n < Q)$ are unicast from the source and the nodes which are greater than $Q(n > Q)$ are multicast. In case of multicasting the node which has the minimum energy per corresponding receiver is set as the relay node. The relay node then forwards the packets from the source to the corresponding receivers.

1) Calculating Residual Energy of a Node: Consider a network with multicast groups $G_1, G_2, \cdots, G_x$. Each group $\{G_i\}$ consists of N nodes. Every node in the MANET calculates its remaining energy periodically. The nodes may operate in either transmission or reception mode. Let $\{E_1, E_2, \cdots, E_n\}$ are the residual energies of the nodes measured by the following method.

The power consumed for transmitting a packet is given by (1)

$$\text{Consumed energy} = TP * t \quad (1)$$

Where TP is the transmitting power and t is transmission time.

The power consumed for receiving a packet is given by (2)

$$\text{Consumed energy} = RP * t \quad (2)$$

Where RP is the reception power and t is the reception time. The value t can be calculated as

$$t = D_s / D_r \quad (3)$$

D_s is Data size and D_r is Data rate

Hence, the residual energy (E) of each node can be calculated using (1) or (2) and (3)

$$E = \text{Current energy} - \text{Consumed energy}$$

2) Algorithm:

1. Consider a group $G_j = \{N_1, N_2, \cdots, N_3\}$
2. Measure the distance d of each node from source S

   $d(S, N_i)$ where $i = 1,2,\cdots,n$
3. Sort the nodes N_i in ascending order of d.
4. Create the partitions $X1$ and $X2$ of the nodes N_i such that

   $$X1 = \{N_1, \cdots, N_Q\}$$

   $$X2 = \{N_{Q+1}, \cdots, N_n\}$$

   Where Q is the distance threshold.
5. Source unicast the packets to $X1$
6. In $X2$ find a relay node N_r which has max (E_i)
7. Then S unicast the packets to N_r which in turn multicast the packets to the rest of the nodes in $X2$.

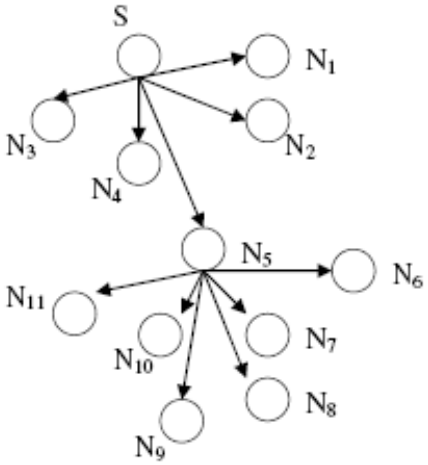


Figure 1. Energy efficient tree construction

Source S unicast the packets to nodes N_1, N_2, N_3, N_4 and N_5 is the relay node. N_5 multicast the packets to the rest of the nodes $N_6, \cdots, N_{11}$.

B. Multicast Admission Control

Most of the existing schemes depend on individual receivers to detect congestion and adjust their receiving rates which are much disadvantageous. We propose a scheme which adjusts the multicast traffic rate at each bottleneck of a multicast tree. Each node estimates its current traffic load and arrival rate. Based on its traffic load, it estimates the receiving rate. If the receiving rate is less than the arrival rate, it adaptively adjusts its receiving rate.

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

In order to adjust the total number of multicast flows which traverse a bottleneck, the following procedure is used. In our proposed scheme, based on the link's output queue state, multicast flows at a bottleneck can be blocked or released. Let the number of packets in the queue is N. Let $QT1$ and $QT2$ $(QT1 < QT2)$ are two thresholds for the queue size. Then the flow is released or blocked based on the following conditions.

If $N \le QT1$, then the multicast flow is released.

If $N > QT2$, then the multicast flow is blocked.

In most of the existing schemes, in order to detect congestion and for adjusting the receiving rate they depend on the individual receivers. In our proposed scheme multicast traffic rate is adjusted at each bottleneck of a multicast tree. Whenever congestion happens or about to, then the multicast sessions which traverse the branch are blocked. Thus the packets are stopped from entering the branch. The blocked flows are released to traverse the branch when the branch is lightly utilized.

C. Multicast Traffic Rate Adjustment

When the available bandwidth is less than the required bandwidth or the queue size is less than a minimum threshold value, it indicates the possibility of congestion or packet loss. The behaviour of the multicast session is expressed as

$$R(t+1) = \begin{cases} R(t) - g & \text{If } R(t) > B \\ R(t) + g & \text{If } R(t) \le B \\ R(t) & otherwise \end{cases}$$

Here $R(t)$ denotes the instantaneous rate of the multicast session at time t. B is the bottleneck bandwidth.

When $R(t) > B$ then the network is congested and the multicast session decreases its rate by a step g.

If $R(t) \le B$ then the network is not congested and the multicast session increases its rate by a step g.

The proposed scheme overcomes most of the disadvantages of existing schemes:

1. Link errors cannot cause the proposed scheme to wrongly block a layer, because instead of the loss information at receivers, the queue state at a bottleneck is used as the metric to adjust the multicast traffic rate at the bottleneck.

2. Link access delay caused by competition in MANETs cannot hinder the rate adjustment in this scheme, because, it blocks multicast layers right at each bottleneck of a multicast tree instead of depending on receivers to request pruning to drop layers.

3. Because of the on-the-spot information collection and rate control this scheme has very limited control traffic overhead.

Moreover, the proposed scheme does not impose any significant changes on the queuing, scheduling or forwarding policies of existing networks.

IV. Simulation Results

A. Simulation Model and Parameters

We use NS2 to simulate our proposed protocol. In our simulation, the channel capacity of mobile hosts is set to the same value: 2 Mbps. We use the distributed coordination function (DCF) of IEEE 802.11 for wireless LANs as the MAC layer protocol. It has the functionality to notify the network layer about link breakage.

In our simulation, 50 mobile nodes move in a 1000 meter x 1000 meter region for 50 seconds simulation time. We assume each node moves independently with the same average speed. All nodes have the same transmission range of 250 meters. In our simulation, the minimal speed is 5 m/s and maximal speed is 5 m/s. The simulated traffic is Constant Bit Rate (CBR).

Our simulation settings and parameters are summarized in table I

TABLE I. SIMULATION PARAMETERS

No. of Nodes	50
Area Size	1000 X 1000
Mac	802.11
Radio Range	250m
Simulation Time	50 sec
Traffic Source	CBR
Packet Size	250,500,…1000
Mobility Model	Random Way Point
Speed	5m/s
Receivers	5,10,…25
Pause time	5 s
Transmit Power	0.660 w
Receiving Power	0.395 w
Idle Power	0.335 w
Initial Energy	3.1 J

B. Performance Metrics

We compare our EERCCP protocol with the multicast AODV [14] protocol. We evaluate mainly the performance according to the following metrics.

Average end-to-end delay: The end-to-end-delay is averaged over all surviving data packets from the sources to the destinations.

Average Packet Delivery Ratio: It is the ratio of the No. of packets received successfully and the total no. of packets sent.

Average Energy Consumption: The average energy consumed by the nodes in receiving and sending the packets are measured.

Throughput: It is the number of packets received by all the nodes in the network.

C. Results

1) Based On Receivers: In this experiment, we vary the group size or the number of receivers per group as 5,10…..25.

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

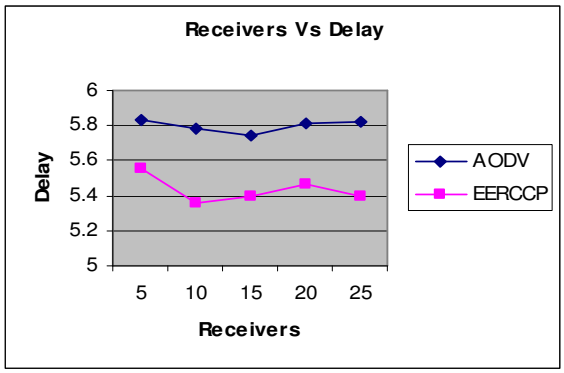


Figure 2. Receivers Vs Delay

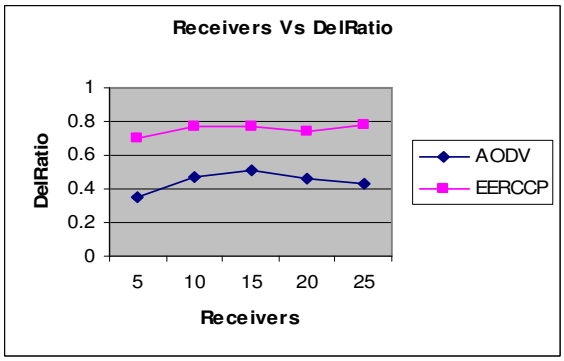


Figure 3. Receivers Vs Delivery Ratio

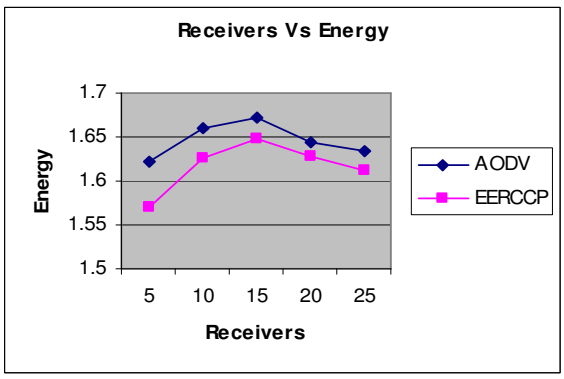


Figure 4. Receivers Vs Energy

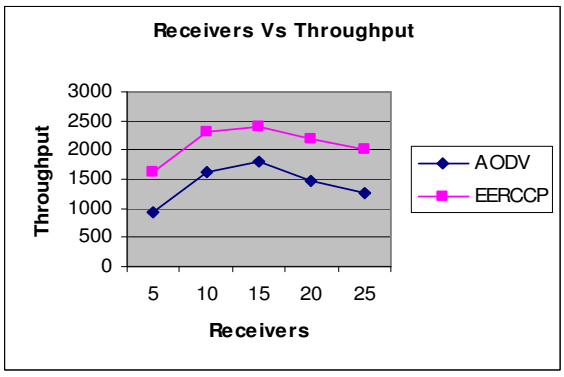


Figure 5. Receivers Vs Throughput

When the number of receivers is increased,

Figure 2 shows the end-to-end delay occurred for both AODV and EERCCP. As we can see from the figure, the delay is less for EERCCP, when compared to AODV.

Figure 3 shows the delivery ratio for both AODV and EERCCP. As we can see from the figure, the delivery ratio is high for EERCCP, when compared to AODV.

Figure 4 shows the energy consumption for both the cases. As we can see from the figure, the energy consumption is less for EERCCP, when compared to AODV.

Figure 3 shows the throughput occurred for both the cases. As we can see from the figure, the throughput is high for EERCCP, when compared to AODV.

2) Based on Psize: In this experiment, we vary the packet size as 250,500…..1000.

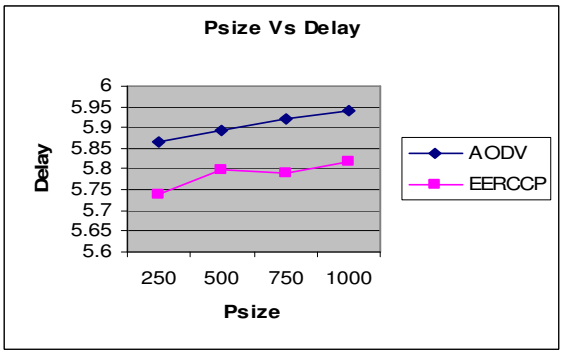


Figure 6. Psize Vs Delay

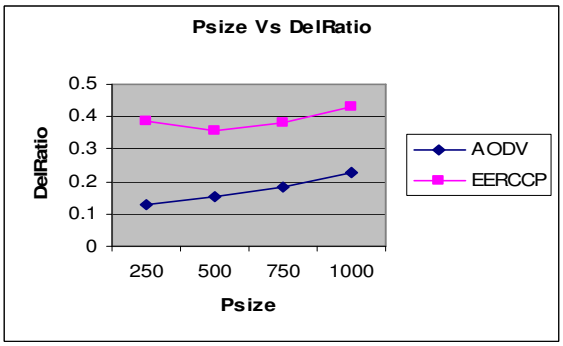


Figure 7. Psize Vs DelRatio

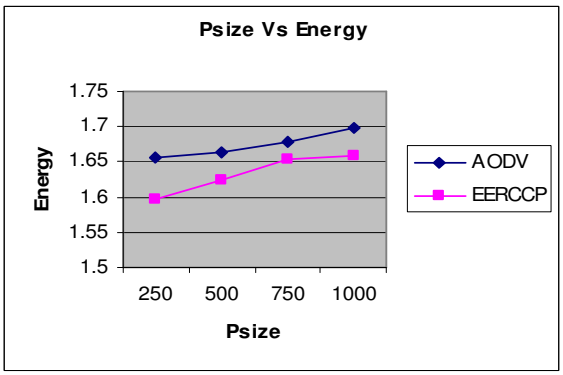


Figure 8. Psize Vs Energy

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

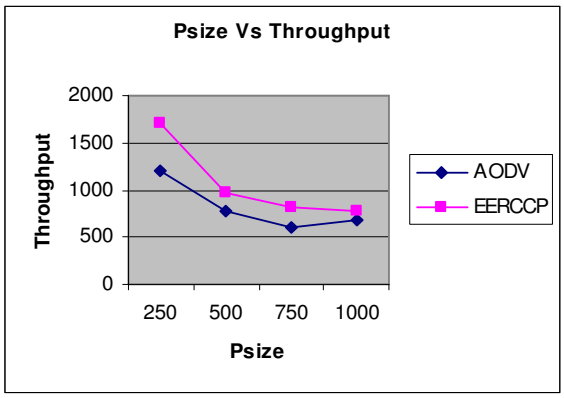


Figure 9. Psize Vs Throughput

When the Psize is increased,

Figure 6 shows the end-to-end delay occurred for both AODV and EERCCP. As we can see from the figure, the delay is less for EERCCP, when compared to AODV.

Figure 7 shows the delivery ratio for both AODV and EERCCP. As we can see from the figure, the delivery ratio is high for EERCCP, when compared to AODV.

Figure 8 shows the energy consumption for both the cases. As we can see from the figure, the energy consumption is less for EERCCP, when compared to AODV.

Figure 9 shows the throughput occurred for both the cases. As we can see from the figure, the throughput is high for EERCCP, when compared to AODV.

V. Conclusion

In this paper, we have proposed an energy efficient and reliable congestion control protocol for multicasting in mobile adhoc networks. Our proposed protocol overcomes the disadvantages of existing multicast congestion control protocols which depend on individual receivers to detect congestion and adjust their receiving rates. In the first phase of our protocol, we have built a multicast tree routed at the source, by including the nodes with higher residual energy towards the receivers. In the second phase, we have proposed an admission control scheme in which a multicast flow is admitted or rejected depending upon on the output queue size. In the third phase, we have proposed a scheme which adjusts the multicast traffic rate at each bottleneck of a multicast tree. Because of the on-the-spot information collection and rate control, this scheme has very limited control traffic overhead and delay. Moreover, the proposed scheme does not impose any significant changes on the queuing, scheduling or forwarding policies of existing networks. Simulation results have shown that our proposed protocol has better delivery ratio and throughput with less delay and energy consumption when compared with existing protocol.

References

[1] Luo Junhai, Xue Liu, Ye Danxia," Research on multicast routing protocols for mobile ad-hoc networks", Elsevier, 2007

[2] Abdussalam Nuri Baryun, and Khalid Al-Begain," A Design Approach for MANET Multicast Protocols", ISBN, 2008

[3] Hua Chen, Baolin Sun," An Entropy-Based Fuzzy Controllers QoS Routing Algorithm in MANET",IEEE,2009

[4] Tolga Numanoglu and Wendi Heinzelman," Improving QoS in Multicasting Through Adaptive Redundancy", University of Rochester Center for Electronic Imaging Systems

[5] D. Agrawal, T. Bheemarjuna Reddy, and C. Siva Ram Murthy," Robust Demand-Driven Video Multicast over Ad hoc Wireless Networks",IEEE,2006

[6] Guojun Wang, Jiannong Cao, Lifan Zhang, Keith C. C. Chan," A Novel QoS Multicast Model in Mobile Ad Hoc Networks", IEEE International Parallel and Distributed Processing Symposium,2005

[7] Vida Lashkari B. O., Mehdi Dehghan" QoS-aware Multicast Ad hoc On-Demand Distance Vector Routing", WCE 2007

[8] Zeyad M. Alfawaer, GuiWei Hua, and Noraziah Ahmed," A Novel Multicast Routing Protocol for Mobile Ad Hoc Networks", ISSN, 2007

[9] Harald Tebbe and Andreas J. Kassler ," QAMNet: Providing Quality of Service to Ad-hoc Multicast Enabled Networks", Wireless Pervasive Computing, 2006 1st International Symposium on, IEEE,2006

[10] Mohammed Saghir, Tat-Chee Wan, Rahmat Budiarto," QoS Multicast Routing Based on Bandwidth Estimation in Mobile Ad Hoc Networks", ICCCE2006

[11] Ravindra Vaishampayan, J.J. Garcia-Luna-Aceves," Efficient and Robust Multicast Routing in Mobile Ad Hoc Networks", IEEE, 2004

[12] Y. Chen and Y. Ko, "A Lantern-Tree Based QoS on Demand Multicast Protocol for A wireless Ad hoc Networks," IEICE Trans. Communications, vol. E87-B, 2004

[13] 13 N. Ben Ali, A. Belghith, J. Moulierac, M. Molnar," QoS multicast aggregation under multiple additive constraints", Elsevier, 2008

[14] Elizabeth M. Royer, Charles E. Perkins," Multicast Operation of the Ad-hoc On-Demand Distance Vector Routing Protocol",ACM,1999

Dr.G.Sasi Bhushana Rao received his BE from GITAM, Visakhapatnam. ME, PhD, MBA (HRD & Marketing) from Osmania University. He has 22 years of research and development, administrative experience as asst. General Manager (CNS) in Airport Authority of India, Ministry of Civil Aviation and ISRO, Govt. of India. Presently he is working as Senior Professor in E C E Dept. at Andhra University Visakhapatnam. His areas of interest are GPS, Signal Processing and Mobile Communications. He has more than 100 in various International and National Journals and conferences (including IEEE, IEE) presently ongoing project works under his guidance are DRDO (N S T L), CSIR, UGC, AICTE and WIPRO research projects in the department of electronics and communication engineering Andhra University Visakhapatnam. He is senior member in IEEE.

M. RajanBabu received B.Tech from Bapatla Engineering College-Bapatla and M.Tech from J N T U College of Engineering-Kakinada in the Department of Electronics And Communication Engineering, He is pursuing his PhD under the guidance of Dr. G. Sasi Bhushana Rao on wireless networks from Andhra University He is having 12 years of teaching experience currently he is working as Associate Prof in Lendi Institute of Engineering and technology Jonnada Vizianagaram (dt) A.P -India

An Intelligent System For Effective Forest Fire Detection Using Spatial Data

K.Angayarkkani
Senior lecturer
Department of ComputerApplications
D.G. Vaishnav College, Arumbakkam, Chennai

Dr.N.Radhakrishnan
Geocare Research Foundation
#23/30, First main Road,
Pammal, Chennai - 600 075, India

Abstract— **The explosive growth of spatial data and extensive utilization of spatial databases emphasize the necessity for the automated discovery of spatial knowledge. In modern times, spatial data mining has emerged as an area of voluminous research. Forest fires are a chief environmental concern, causing economical and ecological damage while endangering human lives across the world. The fast or early detection of forest fires is a vital element for controlling such phenomenon. The application of remote sensing is at present a significant method for forest fires monitoring, particularly in vast and remote areas. Different methods have been presented by researchers for forest fire detection. The motivation behind this research is to obtain beneficial information from images in the forest spatial data and use the same in the determination of regions at the risk of fires by utilizing Image Processing and Artificial Intelligence techniques. This paper presents an intelligent system to detect the presence of forest fires in the forest spatial data using Artificial Neural Networks. The digital images in the forest spatial data are converted from RGB to XYZ color space and then segmented by employing anisotropic diffusion to identify the fire regions. Subsequently, Radial Basis Function Neural Network is employed in the design of the intelligent system, which is trained with the color space values of the segmented fire regions. Extensive experimental assessments on publicly available spatial data illustrated the efficiency of the proposed system in effectively detecting forest fires.**

Keywords- Data Mining, Remote Sensing, Spatial data, Forest Fire Detection, Color Space, Segmentation, Anisotropic diffusion, Radial Basis Function Neural Network (RBFNN).

I. INTRODUCTION

The rapid progress in scientific data collection has led to enormous and ever-increasing quantity of data making it unfeasible to be manually interpreted. Therefore, the development of novel techniques and tools in assist for humans, aiding in the transformation of data into useful knowledge, has been the heart of the comparatively new and interdisciplinary research area called the "Knowledge Discovery in Databases (KDD)" [3]. Data mining is the vital step in KDD, which facilitates the discovery of buried but valuable knowledge from enormous databases. Data Mining is formally defined as "The non-trivial extraction of inherent, new, and potentially valuable information from databases" [5].Data mining combines machine learning, pattern recognition, statistics, databases, and visualization techniques into a single unit so as to enhance efficient information extraction from large databases [6]. Data mining techniques profit a number of fields like marketing, manufacturing, process control, fraud detection and network management. Other than this, a huge variety of data sets like market basket data, web data, DNA data, text data, and spatial data [7] have benefited as well.

The progress in scientific data collection has resulted in huge and continuously rising amount of spatial data [1]. Thus the need, for automated discovery of spatial knowledge from massive amount of spatial data, arises. The process of identifying previously hidden but valuable information from vast spatial databases is known as spatial data mining. It is comparatively tedious to extract patterns of value and interest from the spatial databases owing to the complexity of spatial data types, spatial relationships, and spatial autocorrelation than that of the conventional numeric and categorical data [2]. Spatial data mining technologies facilitate the comprehension of spatial data, discovery of relationships among spatial and non-spatial variables, determination of the spatial distribution patterns of a specific phenomenon further supporting the envisagement of the pattern trends. The elemental parts of spatial data mining are spatial statistics and data mining. Spatial data mining techniques involves visual interpretation and analysis, spatial and attribute query and selection, characterization, generalization and classification, detection of spatial and non spatial association rules, clustering analysis and spatial regression in addition to a wide variety of other fields [8].

Spatial data mining and relational data mining vary from one another because of the fact that in the former the attributes of the neighbors of some object of interest need to be taken into consideration as well, since they have a prominent influence on the object [9]. Some distinguishing characteristics of spatial data that forbid the usage of regular data mining algorithms include: (i) rich data types (e.g., extended spatial objects) (ii) inherent spatial relationships between the variables, (iii) other factors that influence the observations and (iv) spatial autocorrelation among the characteristics [2]. The extraction of patterns of interest and rules from the spatial data sets like the remotely sensed imagery and related ground data significantly benefits the application areas like precision agriculture, community planning, resource discovery and more[10]. Spatial

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

data mining is comprehensively employed in change detection, modeling deforestation, disaster analysis, forest fire detection and other related fields. Our research focuses on the detection of forest fires from the spatial data analogous to forest regions.

A. Forest Fires

For a long time, fires have been a source of trouble. Fires have notable influence over the ecological and economic utilities of the forest, being a prime constituent in a great number of forest ecosystems [11]. Past has witnessed multiple instances of forest and wild land fires. Fires play a remarkable role in determining landscape structure, pattern and eventually the species composition of ecosystems. The integral part of the ecological role of the forest fires [13] is formed by the controlling factors like the plant community development, soil nutrient availability and biological diversity. Fires are considered as a significant environmental issue because they cause prominent economical and ecological damage despite endangering the human lives [12]. Due to the forest fires, several hundred million hectares (ha) of forest and other vegetation are destroyed every year [14].

Occasionally, forest fires have forced the evacuation of susceptible communities in addition to heavy damages amounting to millions of dollars. As per the forest Survey of India 19.27% or 63.3 million ha of the Indian land has been classified as forest area, of which 38 million ha alone are hoarded with resources in great quantity (crown density above 40%).Thus the country's forests face a huge threat. Degradation caused by forest fires [15] jeopardizes the Indian forests. Fires caused huge damage in the year 2007 affecting huge territories in addition to the prominent number of human casualties [16]. Forest fires remains to be a potential threat to ecological systems, infrastructure and human lives. The practical and effective option to minimize the damage caused by the forest fire is to detect the fires at their early stages and reacting fast to prevent the spread of the fire. Hereafter, hefty efforts have been taken to ease the early detection of forest fires, usually being carried out with the help of human surveillance. Forest fire, Drought, Flood and many other phenomena especially the ones with large spatial extent are some of the spatial phenomena which contain predictable spatial patterns that are evident through remote sensing Images/products.

B. Our Contributions

In our earlier work [45], we have presented an efficient forest fire detection system using Fuzzy logic. The primary intention of this research is to extract valuable information from spatial data and employ them for locating the regions vulnerable to forest fire with the aid of Image Processing and Artificial Intelligence techniques. This paper presents an intelligent system that is capable of detecting forest fires. The presented intelligent system utilizes the images in the spatial data that corresponds to forest regions, obtained from remote sensing. The Radial Basis Function Neural Network is employed in the design of the presented intelligent system. The images in the forest spatial data with the presence of fires are utilized in training the neural network. Initially, the digital images in the forest spatial data are converted from RGB to XYZ color space. Then, the segmentation of the image in XYZ color space is carried out with the aid of the renowned anisotropic diffusion approach. The XYZ color space values of the regions with fires, resulting from the segmentation, are fed as input to training the radial basis function neural network. For a given XYZ color space value of a pixel, the trained radial basis function neural network will identify whether that pixel corresponds to fire region or not. The presented intelligent system effectively detects forest fire, which is very well illustrated by the experimental evaluation on the publicly available spatial data.

The rest of the paper is organized as follows: Section II presents a brief review of some recent researches existing in the literature related to forest fire detection. A concise description of the concepts utilized in the presented intelligent system is given in Section III. The proposed intelligent system for effective forest fire detection is presented in Section IV. The experimental results are given in Section V. The conclusions are summed up in Section VI.

II. Review Of Related Researches

The proposed research has been motivated by several earlier researches in the literature related to forest fire detection using spatial data and artificial intelligence techniques. A concise description of some of the recent researches is given in this section.

Armando et al. [17] have studied on the automatic recognition of smoke signatures in lidar signals attained from very small-scale experimental forest fires using neural-network algorithms. A scheme of multi-sensorial integrated systems for early detection of forest fires has been presented by Ollero et al. [18]. The system presented by the authors uses infrared images, visual images, and data from sensors, maps and models. To facilitate the minimization of perception errors and the improvement in reliability of the detection process, it is necessary for the integration of sensors, territory knowledge and expertise, according to their study.

An improved fire detection algorithm which provides increased sensitivity to smaller, cooler fires as well as a significantly lower false alarm rate has been presented by Louis Giglio et al. [19]. The Theoretical simulation and high-resolution Advanced Space borne Thermal Emission and Reflection Radiometer (ASTER) scenes are employed to establish the performance of their algorithm. Seng Chuan Tay et al. [20] have presented an approach to reduce the false alarms in the hotspots of forest fire regions which uses geographical coordinates of hot spots in forest fire regions for detection of likely fire points. The authors employ clustering and Hough transformation to determine regular patterns in the derived hotspots and classify them as false alarms on the assumption that fires generally do not spread in regular patterns such as straight lines. In this work demonstrate the application of spatial data mining for the reduction of false alarm from the set of hot spots is derived from NOAA images.

A graph based forest fire detection algorithm based on spatial outlier detection methods has been presented by Young Gi Byun et al. [21]. By using the spatial statistics the authors have achieved spatial variation in their algorithm. This

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

algorithm illustrates higher user and producer accuracies, when compared with the MODIS fire product provided by the NASA MODIS science team. The ordinary scatter plot algorithm was proved to be inefficient by the authors because it is insensitive to small fires, while Moran's scatter plot was also weak because of the numerical criterion's absence for spatial variation which requires a more and less high commission error.

An approach to predict forest fires in Slovenia using different data mining techniques has been presented by Daniela Stojanova et al. [22]. The authors have employed the predictive models based on the data from a GIS (Geographical Information System) and the weather prediction model - Aladin and MODIS satellite data. The work examined three different datasets: one for the Kras region, one for Primorska region and one for continental Slovenia. The researchers demonstrated that Bagging and boosting of decision trees offers the best results in terms of accuracy for all three datasets. Yasar Guneri Sahin [23] has proposed a mobile biological sensor system for prior detection of forest fires which utilizes animals as mobile biological sensors. This system is based on the existing animals tracking systems used for the zoological studies. The work illustrates that the combination of these fields may lead to instantaneous development of animal tracking as well as forest fire detection. A number of serious forest fires were detected by the system in the earliest, which reduced their effect and therefore contributes to the reduction of the speed of global warming.

A fully automated method of forest fire detection from TIR satellite images on the basis of random field theory has been presented by Florent Lafarge et al. [24]. The results of the system rely only on the confidence coefficient. The obtained values for the both detection rate and false alarm rate were convincing. The estimation of fire propagation direction presents interesting information associated to the evolution of the fires. In Movaghati et al. [25], the capability of agents to be applied in processing of remote sensing imagery has been studied. An agent based approach for forest fire detection has been presented in this paper. The tests used in MODIS version 4 contextual fire detection algorithms were used by the agents to determine agent behavioral responses. The performance of their algorithm was compared against that of MODIS version 4 contextual fire detection algorithm and ground-based measurements. The results portray a good agreement between the algorithms and field data.

In our earlier work [45], we have presented an efficient system to detect forest fires using spatial data collected from forest. Image Processing and Artificial Intelligence techniques were utilized in the design of the presented system. Anisotropic diffusion and fuzzy logic are employed for segmentation and fire detection processes respectively. The images are converted to YCbCr color space and segmentation is performed. The Cr value of YCbCr color space of fire pixels is utilized in the formation of fuzzy sets and fuzzy rules are derived from the formed fuzzy sets. The publicly available spatial data has been employed in the evaluation process. The fuzzy rules derived using the presented system, have successfully detected the forest fires in the spatial data.

III. Description Of Concepts Utilized In The Presented Intelligent System

The concepts utilized in the presented intelligent system for effective forest fire detection such as color space, anisotropic diffusion segmentation and artificial neural networks are detailed in this section.

A. Color Space Conversion

A color space is defined as a means by which the specification, creation and visualization of colors is performed. A computer screen produces colors based on the varied combinations of red, green and blue phosphor emission required to form a color. Typically color is represented by three coordinates or parameters [26]. The location of the color in the color space is exemplified by these parameters. Color space conversion is defined as the transformation and description of a color from one source to another. Normally, color space conversion is performed while converting an image that is represented in one color space to another color space, with the objective of making the translated image appear as similar as possible to the original. The commonly used color spaces are RGB, CIE XYZ, CIE YUV, CIE L*a*b*, YCbCr and HSV. In the proposed intelligent system, the images in RGB color space are converted to XYZ color space.

1) Cie Xyz Color Space: CIE XYZ color space [27] is one of the first mathematically defined color spaces created by the International Commission on Illumination in 1931. Any color can be generated as a mixture of three other colors or "Tristimuli" and commonly RGB for CRT based systems (TV, computer) or XYZ (fundamental measurements). The XYZ color space is defined such that all visible colors can be represented using only positive values, and, the Y value is luminance. As a result, the colors of the XYZ primaries themselves are invisible [28]. The chromaticity diagram is extremely non-linear, in that a vector of unit magnitude denoting the difference between two chromaticities is not uniformly visible. A 3x3 matrix transform is used to transform the RGB values in a particular set of primaries to and from CIE XYZ. These transformations involve tristimulus values which are a set of three linear-light components that conform to the CIE color-matching functions. CIE XYZ is a special set of tristimulus values. The equations to convert RGB into XYZ color space are as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.72169 \\ 0.019334 & 0.119193 & 0.950227 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

B. Anisotropic Diffusion Segmentation

Image segmentation is defined as a low-level image processing task meant for partitioning an image into identical regions [29]. The segmentation results can be used to identify the regions of interest and objects in the scene that is very beneficial to the subsequent image analysis. Because of the fact that the inherent multi-features not only possess non linear relation independently but also encompass inter-feature

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

dependency between R, G, and B, color image segmentation is more monotonous when compared to the grey image segmentation. In our system, we have employed an anisotropic diffusion approach for the segmentation of images in the spatial data. The segmentation is carried out on the XYZ color space converted image.

Literature offers numerous models of linear and nonlinear diffusion for achieving image smoothing and segmentation. Nonlinear anisotropic diffusion has been one of the commonly used approaches by researchers [30], [31] in their works. The anisotropic diffusion enhances the response of edge detection algorithms by a series of operations namely: smoothing the image interiors to emphasize boundaries for segmentation, eliminating the spurious detail and eradicating noise from images efficiently [34]. The relaxation processes that implement anisotropic diffusion tends to leave out the low frequency objects that are complex to be dispersed without over-processing the image.

Anisotropic diffusion in image processing discretizes the family of continuous partial differential equations, which incorporate both the physical processes of diffusion and the Laplacian. Provided that there are no sinks or sources that exist [32], the following equation formulates the abovementioned process (for any dimension):

$$\frac{\partial}{\partial t}u(\bar{x},t) = div\left(c(\bar{x},t)\nabla u(\bar{x},t)\right) \tag{2}$$

Diffusion strength is controlled by $c(x,t)$. Vector x represents the spatial coordinate(s).The ordering parameter is the variable t. The function $u(x,t)$ is considered as image intensity $I(x,t)$ [33].

C. Artificial Neural Networks (ANN)

Artificial Neural Networks are a branch of the artificial intelligence, developed to reproduce human reason and intelligence. ANN possesses the abilities to recognize patterns, manage data and learn like the brain [35]. The weights and the input-output function (transfer function) that is specified for the units are used to characterize the behavior of an ANN [37]. The most significant pros in using artificial neural networks are solving the very complex problems of conventional technologies, not formulating an algorithmic solution or using the very complex solution [35]. In the presented intelligent system, Radial Basis Function Neural Network is employed and is detailed below.

1) Radial Basis Function Neural Network (RBFNN): In the late 80's, a variant of artificial neural network emerged by the name, Radial Basis Functions. Nevertheless, their roots are well-established in much older pattern recognition techniques for instance potential functions, clustering, functional approximation, spline interpolation and mixture models [39]. Radial Basis Function Neural Network (RBFNN) is based on supervised learning. RBF networks were autonomously proposed by many researchers [40], [41], [42], [43], [44] and are a popular variant to the MultiLayer Perceptron MLP. RBF networks are also excellent at modeling non-linear data and can be trained in one stage rather than using an iterative process as in MLP and also learn the given application speedily. The RBF network has a feed forward structure consisting of a single hidden layer of J locally tuned units, which are fully interconnected to an output layer of L linear units. All hidden units concurrently receive the n-dimensional real valued input vector X (Figure. 1). The prime difference from that of MLP is the absence of hidden-layer weights. The hidden-unit outputs are not computed using the weighted-sum mechanism/sigmoid activation; rather each hidden-unit output Z_j is obtained by closeness of the input X to an n-dimensional parameter vector μ_j associated with the j^{th} hidden unit [4]. The response characteristics of the j^{th} hidden unit $(j = 1,2,.....,J)$ is assumed as,

$$Z_j = K\left(\frac{\|X-\mu_j\|}{\sigma_j^2}\right) \tag{3}$$

Where K is a strictly positive radially symmetric function (kernel) with a unique maximum at its 'centre' μ_j and which drops off rapidly to zero away from the centre. The parameter σ_j is the width of the receptive field in the input space from unit j. This implies that Z_j has an appreciable value only when the distance $\|X-\mu_j\|$ is smaller than the width σ_j. Given an input vector X, the output of the RBF network is the L-dimensional activity vector Y, whose l^{th} component $(l = 1,2,....,L)$ is given by [36],

$$Y_l(X) = \sum_{j=1}^{J} w_{lj} Z_j(X) \tag{4}$$

For $l = 1$, mapping of (3) is similar to a polynomial threshold gate. However, in the RBF network, a choice is made to use radially symmetric kernels as 'hidden units'.

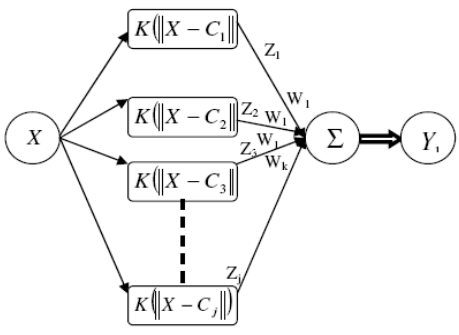


Figure1. Radial Basis Function Neural Network (RBFNN)

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

RBF networks are best suited for approximating continuous or piecewise continuous real-valued mapping $f : R^n \rightarrow R^L$, where n is sufficiently small. These approximation problems include classification problems as a special case. From (3) and (4), the RBF network can be viewed as approximating a desired function $f(X)$ by superposition of non-orthogonal, bell-shaped basis functions. The degree of accuracy of these RBF networks can be controlled by three parameters: the number of basis functions used, their location and their width [38].

IV. Intelligent System For Effective Forest Fire Detection

The proposed intelligent system for effective detection of forest fires is presented in this section. The spatial data collected from forest regions are utilized by the presented intelligent system. With the aid of the images in the spatial data, forest fire detection is performed. The Radial Basis Function Neural Network is employed in the design of the presented intelligent system. The images in the forest spatial data with the presence of fires are employed in training the radial basis function neural network. Initially, the images with the presence of fires are converted from RGB to XYZ color space. The color space conversion from RGB to XYZ is carried out with the help of (1). Figure 2 shows the image in RGB color space and its corresponding XYZ color space converted image.

(a) (b)

Figure 2. a) Image in RGB color space, b) XYZ color space converted image

Afterwards, the XYZ color space converted image is segmented using anisotropic diffusion segmentation, which locates the regions of fire. The result of anisotropic diffusion segmentation is depicted in Figure 3.

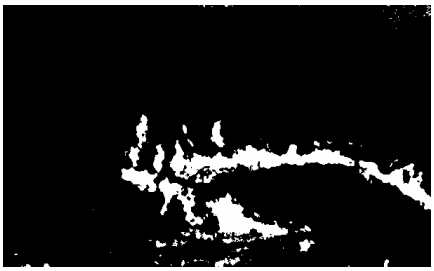
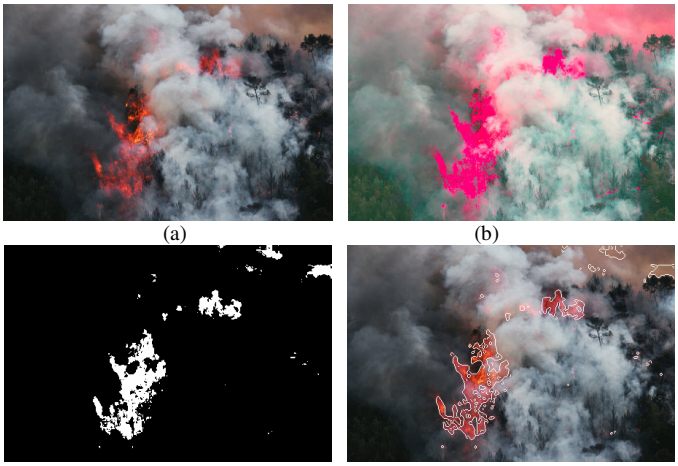


Figure 3. Anisotropic Diffusion Segmented image

The regions of fire obtained using the segmentation is utilized in training the radial basis function neural network. The radial basis function neural network is trained with the XYZ color space values of the pixels that belong to fire regions. With the help of the trained neural network, we can effectively detect the presence of forest fires in an image. The presence of forest fire in an image is detected using the following steps. Initially the image is converted from RGB to XYZ color space. Then, the color space converted image is segmented using anisotropic diffusion segmentation. Subsequently, the XYZ color space values of pixels in the segmented regions are fed as input to the trained neural network for detecting the presence of fires. The designed intelligent system will aid the people in surveillance to detect forest fires and to take appropriate actions.

V. Experimental Results

This section presents the results obtained from the experimentation on the presented intelligent system. The proposed intelligent system is implemented in MATLAB (Matlab 7.4). The publicly available forest spatial data with the presence of fires are employed in training the radial basis function neural network. Consequently, forest spatial data with and without the presence of fires are fed as input to the proposed system for evaluation. The presence of fires is detected effectively by the presented intelligent system with the aid of the trained neural network. The intermediate results of the presented system are depicted in Figure 4. From the results we can conclude that the presented intelligent system can be used for effectively detecting forest fires in the spatial data using artificial intelligence techniques.



(a) (b)

(c) (d)

Figure. 4 Intermediate results of the presented intelligent system a) Input Image, b) XYZ color space converted image, c) Output of Anisotropic Diffusion Segmentation, d) Fire detected Image

VI. Conclusion

Forest fires cause noteworthy environmental demolition while menacing human lives. In the last two decades, a significant effort was made to develop automatic detection tools that could aid the Fire Management Systems (FMS). The three chief trends used for the detection of forest fires are: the use of satellite data, infrared/smoke scanners and local sensors (e.g. meteorological). In this paper, we have presented an intelligent system for effective forest fire detection using spatial data. The proposed system made use of image processing and artificial intelligence techniques. The images in the spatial data, obtained from remote sensing, have been

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

utilized by the presented system for the detection of forest fires. The color space conversion, anisotropic diffusion segmentation and Radial Basis Function Neural Networks have been employed in the presented intelligent system. The experimental results have demonstrated the effectiveness of the proposed intelligent system in detecting forest fires using spatial data.

REFERENCES

[1] Deren Li, Song Xia, Haigang Sui, Xiaodong Zhang, "Change Detection Based On Spatial Data Mining", Wuhan University, white paper, pages-8, 2007.

[2] Shekhar, S., Zhang, P., Huang, Y. and Vatsavai, R.R., Trends in spatial data mining. In: Kargupta, H., Joshi, A. (Eds.), Data Mining: Next Generation Challenges and Future Directions, AAAI/MIT Press, pp. 357-380, 2003.

[3] Martin Ester, Hans-Peter Kriegel, Jörg Sander, "Algorithms and Applications for Spatial Data Mining", Published in Geographic Data Mining and Knowledge Discovery, Research Monographs in GIS, Taylor and Francis, 2001.

[4] Park, J. and Sandberg, I. W., "Universal approximation using radial basis function networks", Neural Computation, vol.3, pp. 246–257, 1991.

[5] W. J. Frawley, P. G. Shapiro, and C. J. Matheus, "Knowledge discovery in databases - an overview," Ai Magazine, vol. 13, pp. 57-70, 1992.

[6] Tadesse, T., J.F. Brown, and M.J. Hayes. 2005. A new approach for predicting drought-related vegetation stress: Integrating satellite, climate, and biophysical data over the U.S. central plains. ISPRS Journal of Photogrammetry and Remote Sensing 59(4):244-253.

[7] K. Julish, Data Mining for Intrusion Detection, a Critical Review, in Applications of Data Mining in Computer Security, D. Barbara and S. Jajodia (Eds.), Kluwer Academic Publisher, 2002.

[8] Hong Tang, and Simon McDonald, "Integrating GIS and spatial data mining technique for target marketing of university courses", ISPRS Commission IV, Symposium 2002, Ottawa Canada, July 9-12 2002.

[9] Imam Mukhlash and Benhard Sitohang, "Spatial Data Preprocessing for Mining Spatial Association Rule with Conventional Association Mining Algorithms," Proceedings of the International Conference on Electrical Engineering and Informatics Institute Teknologi Bandung, Indonesia June 17-19, 2007.

[10] Q.Ding, W. Perrizo, Q.Ding, "On Mining Satellite and Other Remotely Sensed Images," DMKD-2001, pp. 33-40, Santa Barbara, CA, 2001.

[11] González, J.R., Palahí, M., Trasobares, A., Pukkala, T., "A fire probability model for forest stands in Catalonia (north-east Spain)," Annals of Forest Science, 63: 169-176, 2006.

[12] Cortez, P. and A. Morais, "A data mining approach to predict forest fires using meteorological data.", New Trends in Artificial Intelligence, Proceedings of the 13th EPIA - Portuguese Conference on Artificial Intelligence, pp:512--523, 2007.

[13] P.S. Roy , "Forest Fire and Degradation Assessment Using Satellite Remote Sensing and geographic Information System", Proceedings of a Training Workshop Satellite Remote Sensing and GIS Applications in Agricultural Meteorology, pages 361-400, 2003.

[14] Y. Rauste, "Forest Fire Detection with Satellites for Forest Fire Control," Proc. XVIII Congress of ISPRS, Int'l Soc. for Photogrammetry and Remote Sensing, Vol.31, No. B7, pp. 584-588, 1996.

[15] Bahuguna, V.K. & Singh, S., "The forest fire situation in India", Int. Forest Fire News, no. 26, pp. 23-27, 2001.

[16] European-Commission. "Forest Fires in Europe 2007", Technical report, Report No-8, 2008.

[17] Armando M. Fernandes, Andrei B. Utkin, Alexander V. Lavrov, Rui M. Vilar, "Neural Network Based Recognition of Smoke Signatures from Lidar Signals", Neural Processing Letters, Vol. 19, No. 3, pp. 175-189, June 2004.

[18] Ollero, J.R. Martinez-De Dios and B.C. Arrúe, "Integrated systems for early forest-fire detection", III International Conference on Forest Fire Research 14th Conference on Fire and Forest Meteorology VOL II, pp 1977-1988, Luso, 16/20 November 1998

[19] Louis Giglio, Jacques Descloitres, Christopher O. Justice, Yoram J. Kaufman, "An Enhanced Contextual Fire Detection Algorithm for MODIS", Remote Sensing of Environment, vol. 87, pp. 273–282, 2003.

[20] Seng Chuan TAY, Wynne HSU, Kim Hwa LIM, "Spatial Data Mining: Clustering of Hot Spots and Pattern Recognition", Proceedings. 2003 IEEE International Geoscience and Remote Sensing Symposium, Volume: 6, pp: 3685- 3687, 21-25 July 2003.

[21] Young Gi Byun, Yong Huh, Kiyun Yu, Yong Il Kim, "Evaluation of Graph-based Analysis for Forest Fire Detections", Proceedings of world academy of science, engineering and technology, volume. 10, December 2005, ISSN 1307-6884.

[22] Daniela Stojanova, Panče Panov, Andrej Kobler, Sašo Džeroski, Katerina Taškova, "Learning to predict forest fires with different Data Mining techniques", Conference on Data Mining and Data Warehouses (SiKDD 2006),Ljubljana, Slovenia, pp. 255-258, October 9, 2006.

[23] Yasar Guneri Sahin, "Animals as Mobile Biological Sensors for Forest Fire Detection", Sensors, vol. 7, pp. 3084-3099, 2007.

[24] Florent Lafarge, Xavier Descombes, Josiane Zerubia, "Forest fire detection based on Gaussian field analysis", In Proc. European Signal Processing Conference (EUSIPCO), Poznan, Poland, September 2007.

[25] S. Movaghati, F. Samadzadegan, A. Azizi, "An agent-based algorithm for forest fire detection", The International Archives of The Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS Congress Beijing 2008, Volume. XXXVII, Part. B7, Commission: VII/4, Advanced classification techniques, 2008. ISSN 1682-1750.

[26] Adrian Ford and Alan Roberts, "Color Space Conversions", Technical report, August 11, 1998.

[27] CIE, Commission internationale de l'Eclairage proceedings, 1931. Cambridge University Press, Cambridge, 1932.

[28] Smith, Thomas; Guild, John, "The C.I.E. colorimetric standards and their use", Transactions of the Optical Society, vol. 33, no. 3, pp. 73–134, 1931-32.

[29] R. Duda and P. Hart, "Pattern Classification and Scene Analysis", Bayes Decision Theory. John Wiley & Sons, pp. 10-13, 1973.

[30] P. Perona and J. Malik, "Scale space and edge detection using anisotropic diffusion", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 12, Issue 7, July 1990, pp.629 - 639.

[31] L. Alvarez, P.L. Lions, and J.M. Morel, "Image Selective Smoothing and Edge Detection by Nonlinear Diffusion, II," SIAM Journal of Numerical Analysis, vol. 29, no. 3, pp. 845-866, 1992.

[32] Gerig, R. Kikinis, O. Kubler and F. A. Jolesz, "Nonlinear anisotropic filtering of MRI data", IEEE Transactions on Medical Imaging , vol. 11, Issue 2 , pp. 221-232, June 1992.

[33] Sankar Seramani, Zhou Jiayin, Chan Kap Luk, N. Malmurugan and A. Nagappan, "Denoising of MR Images using Non Linear Anisotropic Diffusion Filtering as a Preprocessing Step", International Journal of BioSciences and Technology, Volume 1, Issue 1,2008, pp.17-21.

[34] G. Sapiro, "Cromaticity Diffusion", Proc. of ICIP 2000, Vancouver, BC, Canada, Sept. 2000, vol. II, pp. 784-787.

[35] P. A. Maiellaro, R. Cozzolongo, P. Marino, "Artificial Neural Networks for the Prediction of Response to Interferon Plus Ribavirin Treatment in Patients with Chronic Hepatitis C", Current Pharmaceutical Design, vol. 10, issue. 17, pp. 2101-2109(9), July 2004.

[36] P. Venkatesan and S. Anitha, "Application of a radial basis function neural network for diagnosis of diabetes mellitus", Current Science, vol. 91, no. 9, pp. 1195 - 1199, 10 November 2006.

[37] J.G. Wolff, "Medical Diagnosis as Pattern Recognition in a Framework of Information Compression by Multiple Alignment, Unification and Search", Decision Support Systems, Volume 42, Issue 2, pp: 608 – 625, November 2006.

[38] Park, J. and Sandberg, I. W., "Approximation and radial basis function networks", Neural Computation, vol. 5, pp. 305–316, 1993.

[39] Tou, J.T., Gonzalez, R. C., "Pattern Recognition Principles", Reading, MA: Addison – Wesley, 1974.

[40] Broomhead, D. S. and Lowe, D., "Multivariate functional interpolation and adaptive networks", Complex Systems, vol. 2, pp. 321–355, 1988.

[41] Niranjan, M. A., Robinson, A. J. and Fallside, F., "Pattern recognition with potential functions in the context of neural networks", Proceedings Sixth Scandinavian Conference on Image Analysis, Oulu, Finland, vol. 1, pp. 96–103, 1989.

[42] Moody, J. and Darken, C. J., "Fast learning in networks of locally tuned processing units", Neural Computation, Vol. 1, issue 2, pp. 281–294, 1989.

[43] Hanson, S. J. and Burr, D. J., "Minkowski-r back propagation: learning in connectionist models with non-Euclidean error signals", Neural Information Processing Systems, American Institute of Physics, New York, pp. 348–357, 1988.

[44] Poggio, T. and Girosi, F., "Regularization algorithms for learning that are equivalent to multilayer networks", Science, vol. 247, pp. 978–982, 1990.

[45] K. Angayarkkani and N. Radhakrishnan, "Efficient Forest Fire Detection System: A Spatial Data Mining and Image Processing Based Approach", International Journal of Computer Science and Network Security (IJCSNS), Vol.9 No.3, pp. 100 -107, March 2009.

The author is a postgraduate in Computer Science followed by Master of Philosophy in Computer Science. The author has thirteen years of teaching experience in various fields of computer science. She has enrolled in Mother Teresa Women's University Kodaikanal for her Ph.D. doctoral degree. The author is currently doing research work on spatial data mining and image processing based techniques.

The co-author is a post-graduate in Applied Geology (1985) followed by **M.Tech** degree in **Remote Sensing** (1990) and have completed the **Ph.D doctoral degree** in 1999 on **spatial techniques -** Remote sensing, GPS and GIS - **watershed environment**. The co-author, to his credit, has **Eighteen years** of research and field **experience** in spatial data and geoinformatics - Remote sensing, GIS and GPS - applications. To his credit, he has published twelve research papers in refereed journals – national and international – and international conferences and two papers are under peer review. He has also had the distinct honor of acting as Chairperson for a session on "Ecosystem and Bio-diversity" in an International conference held at Tsukuba University, Ibaraki, Japan, apart from participating many national level seminars, workshops and training programs. He has also been involved in consultancy service to UNESCO, New Delhi, and developed Computer Based Learning tutorial on Geology using VB as front end tool with various graphic utilities (A/V Support) under ICT program. He has been appointed as Lesson writer for M.Sc Geoinformatics covering Satellite Remote sensing and GIS by Annamalai University and has been acting as Resource person for Academic Institutions.

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

Performance Analysis and Optimization of Lumped Parameters of Electrostatic Actuators for Optical MEMS Switches

D.Mohana Geetha
Department of Electronics and Communication Engineering
Kumaraguru college of Technology
Coimbatore641006, India.

M.Madheswaran
Center for Advanced Research
Department of Electronics and Communication Engineering
Muthayammal Engineering College,
Rasipuram 637408 , India.

Abstract – This Paper deals design and simulation of electro statically actuated clamped-clamped beam and cantilever beam using finite element analysis method (FEM). A detailed study and performance analysis for various bias voltages is provided in this paper. The pull in voltages for different dimensions of the beams and the natural or the dominant modes and the corresponding eigen values have been studied for different bias voltages. The displacement of the beam is also studied for various dimensions of the beam. The results were obtained for the length and the width of both clamped – clamped beam and cantilever beam through extensive simulations. The results obtained shows that pull in voltages varies from 2348V to 772V and the natural frequencies vary from 102.92 KHz to 916.35 KHz.

Keywords- electrostatic,optical MEMS,clamped,cantilever,pull-in voltage.

I. I. INTRODUCTION

The development in the field of MEMS and its application in optical domain have increased the attention of research in the recent past.MEMS are found suitable for optical applications because these devices can be matched to optical wavelengths, and manufactured in high volume and high density arrays in the semiconductor manufacturing processes [1]. The inherent advantages of MEMS have started replacing the optical transmitting switches and tunable filters with MEMS actuators [2].The most common process in all MEMS devices is actuation which affect the mechanical motion, forces, and work by a device or system on its surroundings in response to the application of a bias voltage or current. The most common types of actuators are electrostatic, thermal, magnetic, piezoelectric, shape memory alloys, and hydraulics. MEMS based devices are highly used for several applications like biomedical sensors, miniature biomedical instruments ,cardiac management systems, neuro stimulation ,engine and propulsion control, automotive safety, braking and suspension systems, telecommunication optical fiber components and switches, data storage systems, electromechanical signal processing and also for military applications. The drive mechanism of these devices includes a constant voltage source (voltage drive) or constant current source (current drive) to enable electrostatic actuation or capacitive sensing. Sazzadur Chowdhury, M. Ahmadi, W. C. Miller have demonstrated that the electrostatic MEMS devices can be implemented with less complexity[3]. Optical MEMS devices is an electro statically actuated micromechanical mirror [1].

Jin Cheng, Jiang Zhe, Xingtao Wu, K.R.Farmer, V.Modi, Lu Frechette,2004 [4] carried out the pull in analysis for cantilever beam and fixed- fixed beam and have calculated the travel ranges of free deformable actuators to 47.2% and 42% respectively. Ofir Bochobza –Degani, Eran Socher,Yael Nemirovsky [5] had modeled a general electrostatic actuator with a charge distribution in the dielectric coating and results were obtained for pull in voltages. Joseph I.Seeger and Bernhard E.boser [6] demonstrated and showed that the structure can move beyond the well known pull in limit at resonance Sazzadur Chowdhury, M. Ahmadi, W. C. Miller provided the comparative study of closed form methods for calculating the pull in voltage of electro statically actuated fixed- fixed beam actuators[3].

The electrostatic force associated with the constant voltage drive mode becomes nonlinear and gives rise to the well-known phenomenon of 'pull-in'. The pull-in phenomenon causes an electro statically actuated beam to collapse on the ground plane if the drive voltage exceeds certain limit depending on the device geometry. Accurate determination of the pull-in voltage is critical in the design process to determine the sensitivity, frequency response, instability, distortion, and the dynamic range of the device [3]. Since the determination of accurate pull in voltage and natural frequency is essential, extensive simulation is carried out by varying the dimensions of clamped –clamped beam and cantilever beam.

II. MODELING OF ELECTROSTATIC ACTUATORS

Electrostatic actuators have fast response and low power consumption. Application and release of forces take virtually the same time which is not the case of thermo actuation because of fast heating and slow cooling. Electro statically driven actuators are less sensitive to environmental conditions than others. They are two metal structures separated by an air gap. A bias voltage is applied between the metal structures, which results in a separation of charges between them. This produces an electrostatic force that can be used to decrease the gap between the plates as shown in the Fig 1.

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

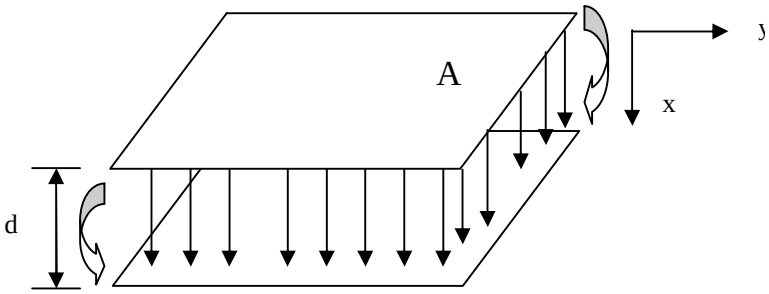


Figure.1.Schmatic structure of parallel plate capacitor

The two plates have an overlapping area of A and a spacing of d. The dielectric constant or relative electrical permittivity of the media between the two plates is denoted ε_r .The permittivity of the media is $\varepsilon = \varepsilon_r \varepsilon_o$,where ε_o is the permittivity of the vacuum[7].

The value of the capacitance C, between the two parallel plates is given by

$$C = \frac{Q}{V} \tag{1}$$

where Q is the amount of stored charge and V is the electrostatic potential.

The electric energy stored in the capacitor is given by

$$U = \frac{1}{2}CV^2 = \frac{1}{2}\frac{Q}{C}^2 \tag{2}$$

According to Gauss's law, the magnitude of the primary electric field E is given by

$$E = \frac{Q}{\varepsilon \mathrm{A}} \tag{3}$$

The magnitude of the voltage is the electric field times the distance between two plates'd'.

The capacitance of the parallel plate capacitor is

$$C = \frac{Q}{V} = \frac{Q}{E.d} = \frac{Q}{\frac{Q}{\varepsilon A}d} = \frac{\varepsilon A}{d} \tag{4}$$

The capacitor can be used as an actuator to generate force or displacement .As a differential voltage is applied between the two parallel plates, an electrostatic attraction force is developed. The magnitude of the forces equals the gradient of the stored electric energy V_s with respect to the dimensional variable .The magnitude of the force is

$$F = \left|\frac{\partial U}{\partial x}\right| = \frac{1}{2}\left|\frac{\partial C}{\partial x}\right|V^2 \tag{5}$$

where x is the dimensional variable.

If the plate moves, the gap between the plates changes and the magnitude of the force can be given as

$$F = \left|\frac{\partial U}{\partial x}\right| = \frac{1}{2}\frac{\varepsilon A}{d^2}V^2 = \frac{1}{2}\frac{CV^2}{d} \tag{6}$$

with normal dimension changed from x to d.

The suspended plate is attracted towards the bottom plate due to the resultant electrostatic force. The suspended plate move towards the bottom plate until an equilibrium exits between them. The suspended plate makes contact with the bottom plate when there is maximum electrostatic force. The capacitance of the device over a range of motion can be used to characterize the electromechanical response of the device.

An electrostatic actuator can be modeled as a variable capacitor suspended by elastic springs. An important design aspect for electrostatic actuators is to determine the amount of static displacement under a certain biased voltage as shown Fig 2. The upper beam is supported by a mechanical spring with a force constant K_m .Gravitational force can be neglected because the mass of the beams are very small.

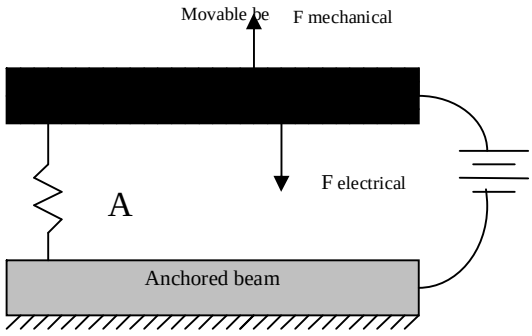


Figure.2.A coupled electromechanical model

When a voltage is applied, an electrostatic force $F_{electric}$ is developed. The magnitude of $F_{electric}$ is given by

$$F_{electric} = \frac{1}{2}\frac{cV^2}{d} \tag{7}$$

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

This force will tend to decrease the gap and gives displacement and mechanical restoring force. Under static equilibrium, the mechanical restoring force has an equal magnitude but opposite direction as the electrostatic force.

The electrostatic force modifies the spring constant and the spring will be softened. The spatial gradient of the electric force is defined as an electrical spring constant and given by

$$K_e = \left|\frac{\partial F_{electric}}{\partial d}\right| = \left|-\left(\frac{cv^2}{d^2}\right)\right| = \frac{cV^2}{d^2}. \quad (8)$$

The electrostatic force at equilibrium when the beam is applied with a bias voltage is given by

$$F_{electric} = \frac{1}{2}\varepsilon A V^2 (x_0 + x)^2 = \frac{\frac{1}{2}c(x)V^2}{x_0 + x} \quad (9)$$

The magnitude of the mechanical restoring force is given by

$$F_{mechanical} = -K_m x \quad (10)$$

Equating the magnitudes of $F_{mechanical}$ and $F_{electric}$ at x and rearranging the terms, the displacement can be calculated as

$$-x = \frac{F_{mechanical}}{K_m} = \frac{F_{electric}}{K_m} = \frac{c(x)V^2}{2(x_{x=0})K_m} \quad (11)$$

At a particular bias voltage, mechanical restoring force and the electrostatic force balance each other. The magnitude of electric force constant equals the mechanical force constant. The effective force constant of the spring is zero. The bias voltage invokes such a condition is called the pull in voltage 'V_p'. If the bias voltage is increased beyond V_p , the equilibrium position disappears. The electrostatic force continues to rise while the mechanical force increases linearly only. The two beams are pulled against each other and they make contact. This condition is called pull in or snaps in.

At the pull in condition the magnitudes of electrical force and mechanical force and can be equated as

$$V^2 = -\frac{2K_m x(x + x_0)^2}{\varepsilon A} = -\frac{2K_m x(x + x_0)^2}{c} \quad (12)$$

The value of x is negative when the spacing between two electrodes decreases. The gradients of these two forces at the intersection point are equal and given by

$$|K_e| = |K_m| \quad (13)$$

From equation (8) and (12)

$$K_e = \frac{cV^2}{(x + x_0)^2} = -2K_m x(x + x_0) \quad (14)$$

The solution of x can be obtained as $x = -\frac{x_0}{3}$ (15)

From the equation (15), it may be concluded that the relative displacement of the parallel plate from its initial position is exactly one third of the original spacing at the critical pull-in voltage.

The voltage at the pull in can be estimated from (12) and equation (15).

The pull in voltage is given by

$$v_p = \frac{2x_0}{3}\sqrt{\frac{K_m}{1.5c}} \quad (16)$$

III. COMPUTATIONAL TECHNIQUE

The clamped-clamped beam and cantilever beam actuators can be simulated and the performance in the electrical domain can be analyzed using Finite element method. The flow graph of the simulation is shown in Fig 3. The structure of the clamped – clamped and cantilever beam actuators are shown in Fig 4.The pull in voltage and the natural frequency of the proposed actuators can be simulated using ANSYS. The region between the two plates can be considered as the active region and the capacitance can be estimated by dividing the beam into meshes. Reduced order modeling of the coupled electrostatic system can be used to replace the electrostatic mesh for improving the execution time.

The eigen value and the eigen vectors can be calculated using

$$[K]\{\Phi_i\} = \lambda_i[M]\{\Phi_i\} \quad (18)$$

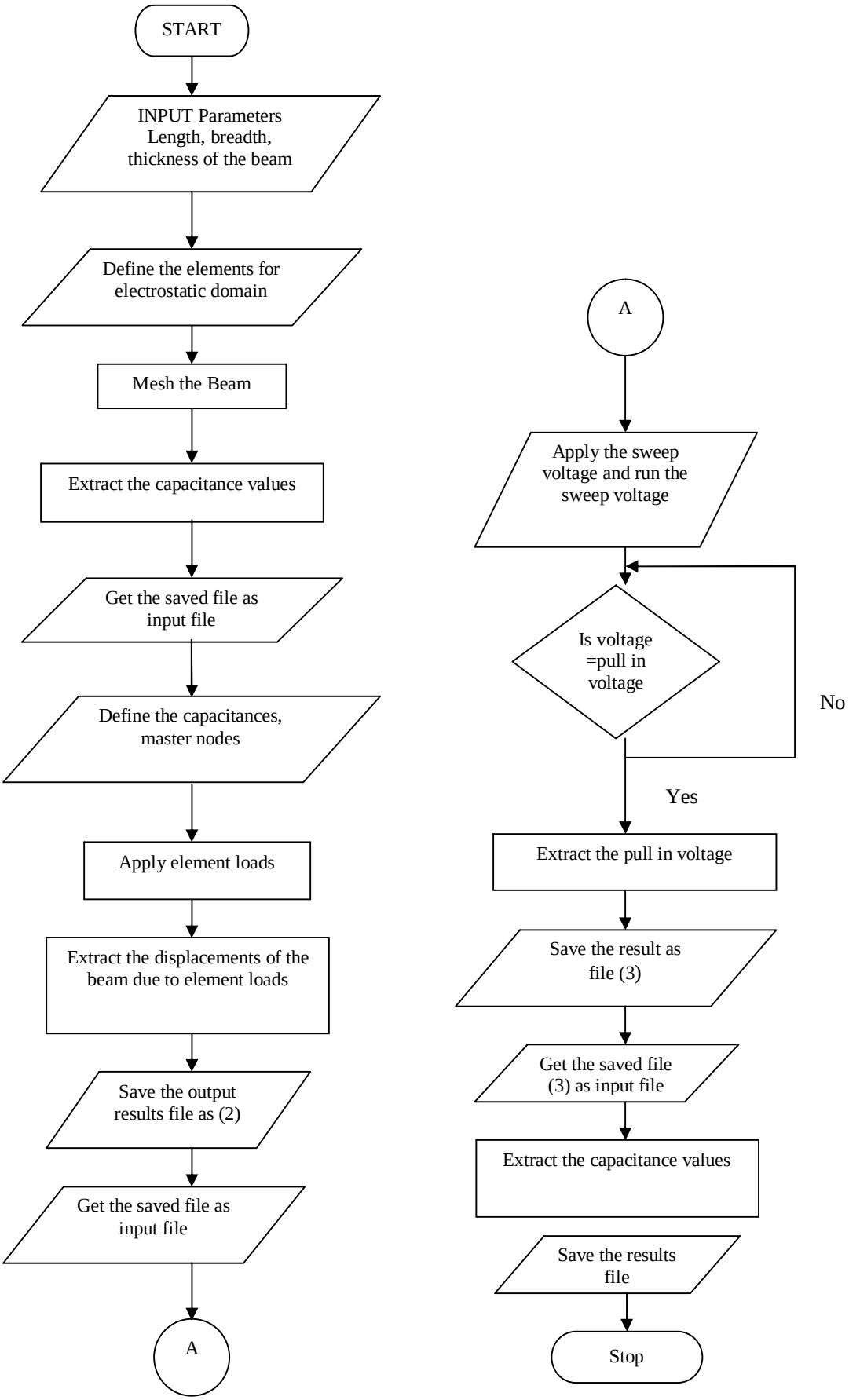
where [K] = structure stiffness matrix

$\{\Phi_i\}$ = eigenvector

λ_i = eigen value

[M] = structure mass matrix

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



Figuer.3.Flow chart for Pull in voltage and extracting capacitance

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

The structure of the stiffness and mass matrix can be reduced by considering the set of degrees of freedoms (DOFs).Then using Guyan reduction technique the relationship can be rewritten as $[k]\{\hat{\phi}_i\} = \lambda_i [\hat{M}]\{\hat{\phi}_i\}$ (19)

Where $[k]$ = reduced stiffness matrix (known)

$\{\hat{\phi}_i\}$ = eigenvector (unknown)

λ_i = eigen value (unknown)

$[\hat{M}]$ = reduced mass matrix (known)

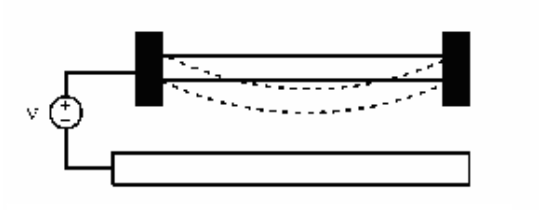


Figure 4(a).Clamped-Clamped beam

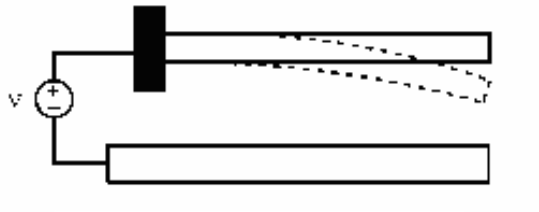


Figure 4(b) Cantilever beam

Then the actual eigen value can be extracted using HBI (Householder-Bisection-Inverse iteration).Using FEM and the ANSYS , the pull in voltage and the natural frequency can be obtained to estimate the performance of both the actuators to find their capacity to use them in optical MEMS switches.

V. RESULTS AND DISCUSSION

The electrostatic actuators with clamped-clamped beam and the cantilever beam structures has been simulated using ANSYS and the parameters have been estimated using FEM analysis.The simulated output of the clamped-clamped beam and cantilever electrostatic actuators is shown in Fig 5(a) and 5(b)respectively. The pull in voltage for various beam length has been calculated for different beam width of clamped-clamped beam actuators and shown in Fig .6

It is seen that the pull in voltage decreases with beam length as well as beam width. However it is clear from the results that the variation in pull in voltage is much significant for the lower beam width. The pull voltage for cantilever beam with various beam width and beam length is shown in Fig.7.It is seen from the Fig that the pull in voltage for the beam length 90 µm and 2 µm beam width is 450V.It is found that at the same beam width and the beam length the pull voltage was estimated as 1100v in clamped-clamped beam actuators. The natural frequency for both clamped –clamped beam and cantilever beam actuators are shown in Table.I

It is seen from the table that the natural frequency is inversely proportional to the beam length in both the cases. The variation of natural frequency with the length and width of the cantilever beam is shown in the Fig .8

The displacement of the beam for length L=85 µm for both clamped –clamped and cantilever beam in Fig.9.It is seen from the Fig that the displacement is much significant in cantilever beam for the beam with less width.

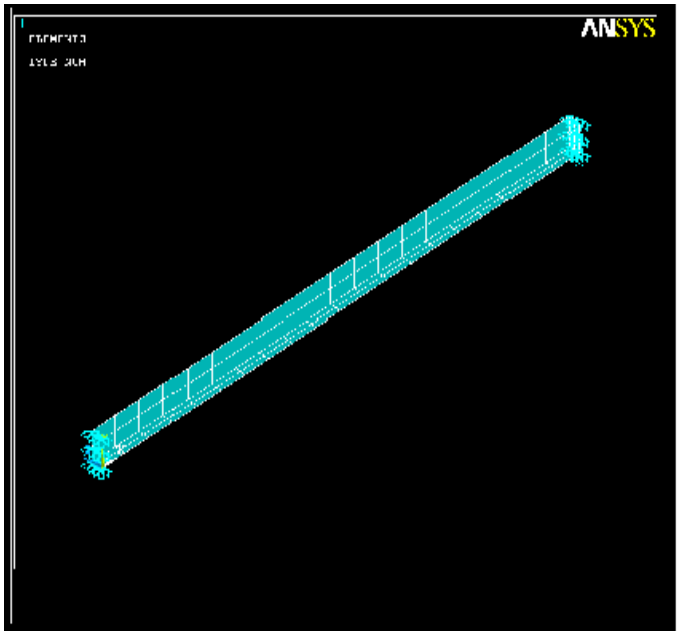


Figure 5(a).Output of displaced clamped-clamped beam.

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

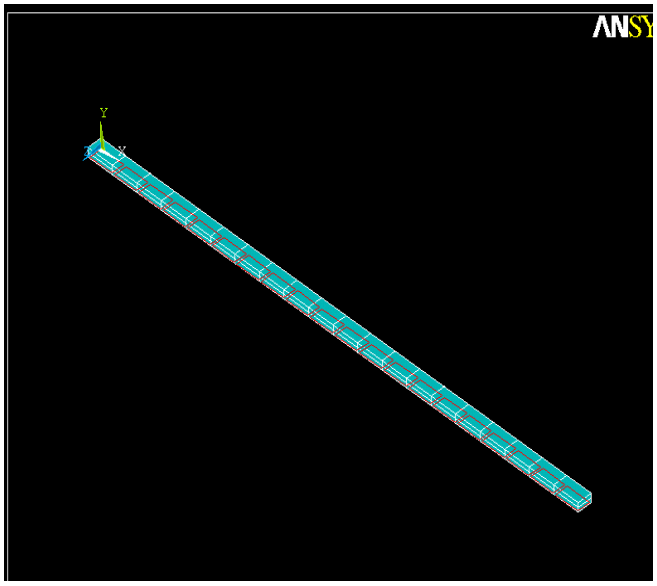


Figure 5(b).Output of displaced cantilever beam.

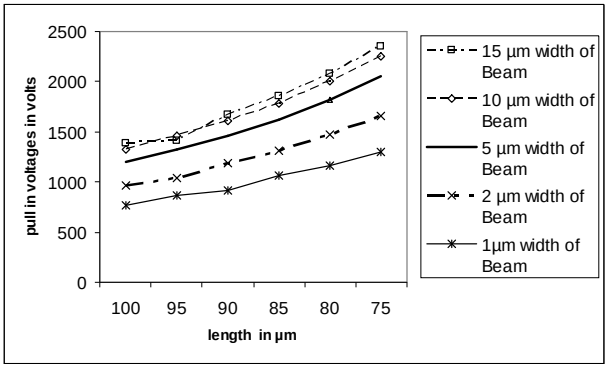


Figure. 6 Pull in voltages of clamped-clamped

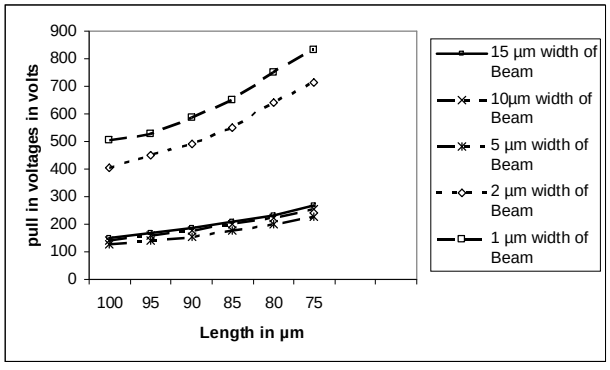


Figure.7 Pull in voltage of cantilever beam.

TABLE I.

NATURAL FREQUENCIES FOR CLAMPED AND CANTILEVER BEAMS

Beam Width in µm	Beam Length in µm	Natural Frequencies (Hz)	
		Clamped-clamped beam	Cantilever beam
15	100	0.10292E+08	0.48965E+07
10	100	0.10289E+08	0.48960E+07
5	100	0.10285E+08	0.48957E+07
2	100	0.10284E+08	0.17325E+07
1	100	0.10284E+08	0.17325E+07
15	95	0.10294E+08	0.54232E+07
10	95	0.11337E+08	0.54226E+07
5	95	0.11333E+08	0.54222E+07
2	95	0.11332E+08	0.19193E+07
1	95	0.11332E+08	0.19193E+07
15	90	0.12566E+08	0.60395E+07
10	90	0.12562E+08	0.60389E+07
5	90	0.12558E+08	0.60384E+07
2	90	0.12556E+08	0.21381E+07
1	90	0.12556E+08	0.21381E+07
15	85	0.14011E+08	0.67670E+07
10	85	0.14006E+08	0.67662E+07
5	85	0.14001E+08	0.67657E+07
2	85	0.14000E+08	0.23965E+07
1	85	0.13999E+08	0.23965E+07
15	80	0.15730E+08	0.76340E+07
10	80	0.15725E+08	0.76330E+07
5	80	0.15719E+08	0.76324E+07
2	80	0.15717E+08	0.27046E+07
1	80	0.15716E+08	0.27046E+07
15	75	0.17797E+08	0.86784E+07
10	75	0.17791E+08	0.86773E+07
5	75	0.91649E+07	0.86765E+07
2	75	0.17781E+08	0.30762E+07
1	75	0.17781E+08	0.30762E+07

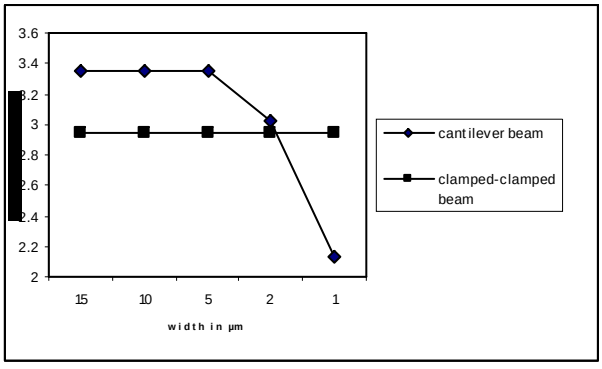


Figure.9Variation of displacement of the beams for length L=85 µm.

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

VI. CONCLUSION

The pull in voltages calculations was made using finite element method. ANSYS with Reduced order modeling was used to extract results quickly. Fringing field effects are also considered to yield accurate results. Even with the strongly coupled lumped transducers, convergence issues were experienced when applied to the difficult hysteric pull-in and release analyses. The cause of the problem can be attributed to the negative total system stiffness matrix and can be resolved using the augmented stiffness method.

REFERENCES

[1] Jinghong Chen, *Member, IEEE*, Wendelin Weingartner, Alexi Azarov, and Randy C. Giles"Tilt-Angle Stabilization of Electro statically Actuated Micromechanical Mirrors Beyond the Pull-In Point" Journal of Microelectromechanical Systems, Vol. 13, No. 6, December 2004.

[2] Marcel W. Pruessner, *Student Member, IEEE*, Kuldeep Amarnath, Madhumita Datta, *Member, IEEE*, Daniel P. Kelly,S. Kanakaraju, Ping-Tong Ho, and Reza Ghodssi, *Member, IEEE "* InP-Based Optical Waveguide MEMS Switches With Evanescent Coupling Mechanism" Journal of Microelectromechanical Systems, Vol. 14, No. 5, October 2005.

[3] Sazzadur Chowdhury, M. Ahmadi, W. C. Miller "A Comparison of Pull-in Voltage Calculation Methods for MEMS-Based Electrostatic Actuator Design"1st International Conference on Sensing Technology, November 21-23, 2005 Palmerston North, New Zealand.

[4] Jin Cheng, Jiang Zhe, Xingtao Wu, K.R.Farmer, V.Modi, Lu Frechette"Analytical and FEM Simulation Pull- in study on Deformable Electrostatic Micro Actuators" Nanotech 2002 vol.1 Technical proceedings of the 2002 International conference on Modeling and simulation of Microsystems.

[5] Ofir Bochobza –Degani, Eran Socher,Yael Nemirovsky" On the effect of residual charges on the pull- in parameters of electrostatic actuators "Sensors and Actuators,2002,Elsevier press.

[6] Joseph I.Seeger and Bernhard E.boser"Parallel Plate driven oscillations and resonant pull in"solid state sensor,actuator and Microsystems workshop,2002.

[7].Chang Liu"foundations of MEMS"Illinois ECE series,Pearson International edition 2006.

AUTHORS PROFILE

Dr. M. Madheswaran has obtained his Ph.D. degree in Electronics Engineering from Institute of Technology, Banaras Hindu University, Varanasi in 1999 and M.E degree in Microwave Engineering from Birla Institute of Technology, Ranchi, India. He has started his teaching profession in the year 1991 to serve his parent Institution Mohd. Sathak Engineering College, Kilakarai where he obtained his Bachelor Degree in ECE. He has served KSR college of Technology from 1999 to 2001 and PSNA College of Engineering and Technology, Dindigul from 2001 to 2006. He has been awarded Young Scientist Fellowship by the Tamil Nadu State Council for Science and Technology and Senior Research Fellowship by Council for Scientific and Industrial Research, New Delhi in the year 1994 and 1996 respectively. His research project entitled "Analysis and simulation of OEIC receivers for tera optical networks" has been funded by the SERC Division, Department of Science and Technology, Ministry of Information Technology under the Fast track proposal for Young Scientist in 2004. He has published 120 research papers in International and National Journals as well as conferences. He has been the IEEE student branch counselor at Mohamed Sathak Engineering College, Kilakarai during 1993-1998 and PSNA College of Engineering and Technology, Dindigul during 2003-2006. He has been awarded Best Citizen of India award in the year 2005 and his name is included in the Marquis Who's Who in Science and Engineering, 2006-2007 which distinguishes him as one of the leading professionals in the world. His field of interest includes semiconductor devices, microwave electronics, optoelectronics and signal processing. He is a member of IEEE, SPIE, IETE, ISTE, VLSI Society of India and Institution of Engineers (India).

Mrs.D.Mohanageetha obtained her B.E degree from Bharathiar University in 1996, M.E Degree in Communication systems from Madurai Kamaraj University, Madurai in 2000. She has started her teaching profession in the year 1996 at V.L.B Janakiammal College of Engineering and Technology, Coimbatore. At present ,she is an Assistant Professor in Department of Electronics and Communication, Kumaraguru college of Technology, Coimbatore, Tamil nadu,India . She has published 10 research papers in International and national conferences. She is a part time Ph.D research scalar in Anna University Chennai. Her areas of interest are Optical networking,, MEMS,EMI/EMC and Image processing. She is a life member of ISTE and member of IEEE.

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

Modeling of Human Criminal Behavior using Probabilistic Networks

Ramesh Kumar Gopala Pillai
Research Scholar
R.V. Center for Cognitive Technologies
Bangalore, India

Dr. Ramakanth Kumar .P
Professor
R.V. Center for Cognitive Technologies
Bangalore, India

Abstract— ***Currently, criminal's profile (CP) is obtained from investigator's or forensic psychologist's interpretation, linking crime scene characteristics and an offender's behavior to his or her characteristics and psychological profile. This paper seeks an efficient and systematic discovery of non-obvious and valuable patterns between variables from a large database of solved cases via a probabilistic network (PN) modeling approach. The PN structure can be used to extract behavioral patterns and to gain insight into what factors influence these behaviors. Thus, when a new case is being investigated and the profile variables are unknown because the offender has yet to be identified, the observed crime scene variables are used to infer the unknown variables based on their connections in the structure and the corresponding numerical (probabilistic) weights. The objective is to produce a more systematic and empirical approach to profiling, and to use the resulting PN model as a decision tool.***

Keywords-component; Modeling, criminal profiling, criminal behavior, probabilistic network, Bayes Rule

I. INTRODUCTION

Modeling human criminal behavior is challenging due to many variables involved and the high degree of uncertainty surrounding a criminal act and the corresponding investigation. Probabilistic graphs are suitable modeling techniques because they are inherently distributed and stochastic. In this paper, the system variables comprising the PN are offender behaviors and crime scene evidence, which are initialized by experts through their professional experience or expert knowledge.

The mathematical relationships naturally embedded in a set of crimes [3, 4, 8] are learned through training from a database containing solved criminal cases. The PN model is to be applied when only the crime scene evidence is known to obtain a useable offender profile to aid law enforcement in the investigations. A criminal profile is predicted with a certain quantitative confidence.

The PN approach presented here seeks to build on the ideas of behavior correlations in order to obtain a usable criminal profile when only crime scene evidence is known from the investigation.

This paper proposes a systematic approach for deriving a multidisciplinary behavioral model of criminal behavior. The proposed offender behavioral model is a mathematical representation of a system comprised of an offender's actions and decisions at a crime scene and the offender's personal characteristics.

The influence of the offender traits and characteristics on the resulting crime scene behaviors is captured by a probabilistic graph or PN that maps cause-and-effect relationships between events, and lends itself to inductive logic for reasoning under uncertainty [1]. The use of PNs for CP may allow investigators to take into consideration various aspects of the crime and discover behavioral patterns that might otherwise remain hidden in the data. The various aspects of a crime include a victimology assessment (victim's characteristics, e.g., background characteristics, age, gender, and education), crime scene analysis (evidence from the crime scene, e.g., time and place where the crime occurred), and a medical report (autopsy report, e.g., type of non-deadly and deadly lesions and signs of self defense).

The PN approach to criminal profiling is demonstrated by learning from a series of crime scene and offender behaviors. The learning techniques employed in this modeling research are evaluated on a set of validation cases not used for training by defining a prediction accuracy based on the most likely value of the output variables (offender profile) and its corresponding confidence level.

II. APPROACH

To start with, a graphical model of offender behavior is learned from a database of solved cases. The resulting CP model obtained through training is then tested by comparing its predictions to the actual offenders' profiles.

Let the database sample space = D,

Let D consist of 'd' solved cases $\{C_1, ...,C_d\}$,

where C_i is an instantiation of X, which is randomly partitioned into two independent datasets such as a training set

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

T and a validation set V, such that D = T *U* V. The variables in X are partitioned as follows: the inputs are the crime scene (CS) variables X^I (evidence) for $X^I = (X^I_1 , ...,X^I_k)$, and the outputs are the offender (OFF) variables comprising the criminal profile X^O, for $X^O = (X^O_1 , ...,X^O_m)$, where $(X^I ,X^O) \varepsilon X$.

The PN model is learned from T, as explained later, and it is tested by performing inference to predict the offender variables (OFF) in the validation cases V. An offender profile is estimated based on crime scene evidence, with a prediction being the most likely value of a particular offender variable. During the testing phase, the predicted value of X^O_i , denoted by $x^P_{i,a}$ where a=1 or 2 for a binary variable, is compared to the observed state $x^O_{i,b}$ obtained from the validation set V, where b=1 or 2. An example of an offender variable is "gender", with states "male" and "female". The overall performance of the PN model is evaluated by comparing the true (observed) states $x^O_{i,b}$ to the predicted output variable values $x^P_{i,a}$ in the validation cases. This process tests the generalization properties of the model by evaluating its efficiency over V.

III. VARIABLES CONSIDERED

The relevant categories of variables that have emerged from the criminal profiling research as selected by investigators, criminologists, and forensic psychologists are described as follows:

• *Crime Scene Analysis* (CSA): CSA variables are systematic observations made at the crime scene by the investigator. Examples of CSA variable pertain to where the body was found (e.g., neighborhood, location, environment characteristics), how the victim was found (e.g., the body was well-hidden, partially hidden, or intentionally placed for discovery), and the correlation between where the crime took place and where the body was found (e.g., the body was transported after the murder).

• *Victimology Analysis* (VA): VA variables consist of the background characteristics of the victim independent of the crime. For example, VA variables include the age, sex, race, education level, and occupation of the victim.

• *Forensic Analysis* (FA): FA variables rely on the medical examiner's report that deals with the autopsy. Examples of this are time of death, cause of death, type of non-lethal wounding, wound localization, and type of weapon that administered the wounds.

The set of CP variables used in this paper were defined in previous research [4, 5, 7, 8]. The selection criteria for variable selection [6] are:

(i) Behaviors are clearly observable and not easily misinterpreted

(ii) Behaviors are reflected in the crime scene, e.g., type of wounding, and

(iii) Behaviors indicate how the offender acted toward and interacted with the victim

e.g., victim was bound/gagged, or tortured. Some crime scene (CS) variables describing the observable crime scene and offender (OFF) variables describing the actual offender were selected based on the above criteria. Examples of the CS variables are multiple wounding to one area, drugging the victim, and sexual assault. Examples of the offender variables include prior offenses, relationship to the victim, prior arrests, etc. The variables all have binary values representing whether the event was present or absent.

IV. TRAINING THE MODEL

The basic schematic of the training software, including the validation process, is shown in Figure 1, where P^h is the proposed PN and P^{opt} is the trained (or optimized) PN. The software is intended to aid law enforcement in the investigation of violent crimes. Because the cases are unsolved and only the crime scene inputs are known, the criminal profiling software consists of a trained PN model that has been previously trained and validated with D. Also, the model has the potential to be updated by means of an incremental training algorithm when additional cases are solved by the police. Thus, $P^{trained}$ consistently reflects the model of an evolving criminal profile over time.

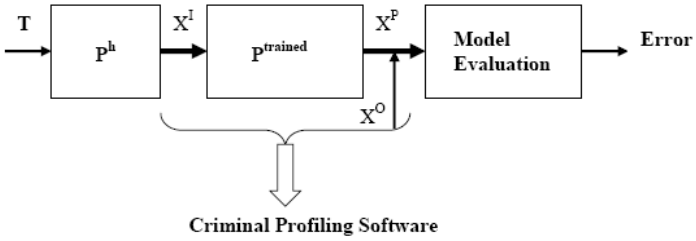


Figure 1: CP model training and validation software

V. DATA BASE OF SOLVED CASES

A set of single offender/single victim homicides was collected by psychologists from solved homicide files of the British police [4, 6]. In order to examine the aggressive behavioral patterns of a particularly violent offense, the criteria for case selection is: single offender/single victim homicide cases; a mixture of domestic (where the victim and offender were known to each other (e.g., family member, spouse, co-worker) and stranger (the offender is unknown to the victim, thus they had no previous links to each other) cases; offenders are adults at least 17 years of age, as defined by the court system. Excluded from the sample were cases when the cause of death was not aggressive or extremely intentional. Homicides by reckless driving are not included due to the lack of interpersonal interaction between the offender and victim.

VI. SAMPLING

A simulation set is built to produce an artificial CP database to study the PN learning and inference capabilities. This included a more extensive list of crime scene, offender characteristics

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

and multiple-valued variables. A PN is used to simulate a set of cases where the crime scene and offender variables can be chosen by the user. An initial structure thus relating the variables and the corresponding initial probabilistic parameters θ_0 are declared based on the prior knowledge, through experience, or by sampled statistics. Cases are simulated by feed forward sampling, where variables are sampled one at a time in order from top-level variables (variables without parents), to the mid-level variables (variables with both parents and children), ending with the bottom-level variables (children variables with parents only). For each variable, the discrete conditional prior probabilities in vector form are given as:

$[P(x_{i,1}|\ \pi_i),\ (x_{i,2}|\ \pi_i),\ldots..,\ (x_{i,ri}|\ \pi_i)]$

where r_i is the maximum state for X_i and π_i disappears if X_i is a top-level variable. A value v_i is drawn from a uniform continuous distribution between '0' and '1' and the conditional prior probability vector as a vector of ranges becomes

$[P(x_{i,1}|\ \pi_i), P(x_{i,1}|\ \pi_i) + P(x_{i,2}|\ \pi_i),\ldots..,$

$\sum_{i=1}^{r_i} (x_{i,j}|\ \pi_i)]$ *which refers to*

$$X_i = \begin{cases} x_{i,1} \text{ if } 0 < v < P(x_{i,1}|\pi_i) \\ x_{i,2} \text{ if } P(x_{i,1}|\pi_i) \le v < \sum_{i=1}^{r_i} P(x_{i,j}|\pi_i) \\ x_{i,ri} \text{ if } \sum_{i=1}^{r_i} P(x_{i,j}|\pi_i) \le v < 1 \end{cases} \quad (1)$$

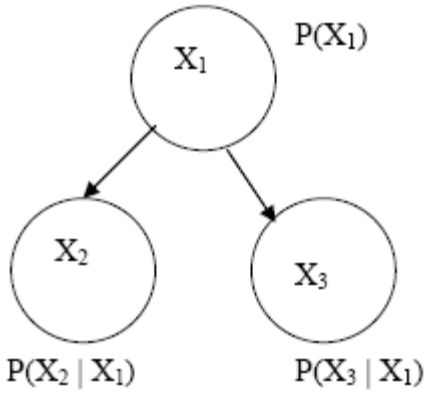


Figure 2: Three-nodal model

To simulate a set of cases for the system represented by the three-nodal model illustrated in Figure 2, the variables are ordered as (X_1, X_2, X_3), where X_1 is the parent of X_2 and X_3, and $X_1 = (x_{1,1}, x_{1,2})$, X2 = ($x_{2,1}$, $x_{2,2}$) and X3 = ($x_{3,1}$, $x_{3,2}$). Starting with X_1, it has three possible states with the prior probabilities $P(x_{1,1}) = 0.2$, $P(x_{1,2}) = 0.5$, and $P(x_{1,3}) = 0.3$, which becomes a vector [0.2, 0.7, 1] referring to:

$$X_1 = \begin{cases} x_{1,1} \text{ if } 0 \le v_1 < 0.2 \\ x_{1,2} \text{ if } 0.2 \le v_1 < 0.9 \\ x_{1,3} \text{ if } 0.9 \le v_1 < 1 \end{cases} \quad (2)$$

If $v_1 = 0.11$ which makes $X_1 = x_{1,1,}$ and the Conditional Probability Table (CPT) for X2 is listed in Table 1.

X_2	$P(x_{2,1}\|\ X_1)$	$P(x_{2,2}\|\ X_1)$
$X_1 = x_{1,1}$	0.2	0.8
$X_1 = x_{1,2}$	0.9	0.1

Table 1 Conditional probability Table

Then the conditional prior probability vector of ranges for a newly generated v_2 becomes

$$X_2 = \begin{cases} x_{2,1} \text{ if } 0 \le v_2 < 0.2 \\ x_{2,2} \text{ if } 0.2 \le v_2 < 1 \end{cases} \quad (3)$$

X_3 is sampled following the same procedure as X_1 and X_2. This is repeated until the desired number of cases as specified by the user is reached. The Matlab function utilized for the sampling exercise is sample b_net in the Bayes Net Toolbox [2].

VII. PN PREDICTIONS AND ACCURACY

When a PN model of offender behavior on the crime scene is learned from solved cases, it is implemented on a set of solved validation cases in order to test the trained model's performance. Performance is tested through probabilistic inference. Inference is the process of updating the probability distribution of a set of possible outcomes based upon the relationships represented by the PN model and the observations of one or more variables. With the updated probabilities, a prediction can be made from the most likely value of each inferred variable. Thus, in order to test the trained model, only the crime scene evidence is inserted into the model, with the predicted offender profile being compared to the actual offender characteristics. Because this is a probabilistic model, a certain confidence accompanies the offender variable predictions.

VIII. CONCLUSIONS

This paper presents an approach for deriving a network model of criminal profiling that draws on knowledge-based systems and on fields of criminology and offender profiling. Implementing probabilistic networks makes it possible to represent multidimensional interdependencies between all relevant variables that have been identified in previous research as playing a role in determining or reflecting the behavior of offenders at the crime scene. Hence, a valid

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

network model can be used to predict unknown variables composing an offender profile based on the variables observed from the crime scene.

REFERENCES

[1] R. Cowell. Introduction to inference for Bayesian networks. In M.I. Jordan, editor, Learning in Graphical Models, pages 9–26. 1998

[2] K. Murphy. How To Use Bayes Net Toolbox. Refer http://www.ai.mit.edu/ murphyk/Software/BNT/bnt.html, 2004.

[3] R.K. Ressler, A. Burgess, and J.E. Douglas. Sexual Homicide: Patterns and motives. New York: Lexington Books, 1988.

[4] C.G. Salfati. Profiling homicide: A multidimensional approach. Homicide Studies,4:265–293, 2000.

[5] C.G. Salfati. Greek homocide, a behavioral examination of offender crime-scene actions. Homicides Studies, 5(4):335–362, November 2001.

[6] C.G. Salfati. Offender interaction with victims in homicide: A multidimensional analysis of crime scene behaviors. Journal of Interpersonal Violence, 18(5):490– 512, 2003.

[7] C.G. Salfati and F. Dupont. Canadian homicide: An investigation of crime scene actions. Homicide Studies, In Press for 2005.

[8] P. Santtila, H. Hakkanen, D. Canter, and T. Elfgren. Classifying homicide offenders and predicting their characteristics from crime scene behavior. Scandinavian Journal of Psychology, 44:107–118, 2003.

AUTHORS PROFILE

Mr. Ramesh Kumar Gopala Pillai is a post graduate in Information Technology from Kuvempu University. He also possesses a MBA degree in Banking and Finance. Currently he is pursuing his research in the area 'Modeling of Human Criminal Behavior and its Implications' under Kuvempu University. He has published a couple of papers in reputed International conferences and journals.

Dr. Ramakanth Kumar P. is a doctorate from Mangalore University. He has successfully guided research scholars to their PhD degree. He has executed several research projects as Principal Investigator for Government and Research establishments like DRDO, ISRO etc. He has several publications in reputed refereed International Journals and Conferences. Currently he is working as a Professor at R V Center for Cognitive Technologies, Bangalore.

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

Reaching the unreached

A Role of ICT in sustainable Rural development.

Mr.Nayak S.K.
Head, Dept. of Computer Science
Bahirji Smarak Mahavidyalaya,
Basmathnagar, Dist.Hingoli. (MS),
India

Dr.S.B.Thorat
Director,
Institute of Technology and Mgmt.
Nanded, Dist.Nanded. (MS),
India

Dr.Kalyankar N.V.
Principal
Yeshwant Mahavidyalaya, Nanded
Nanded (MS)
India

Abstract—

We have seen in last few decades that the progress of information technology with leaps and bounds, which have completely changed the way of life in the developed nations. While internet has changed the established working practice and opened new vistas and provided a platform to connect, this gives the opportunity for collaborative work space that goes beyond the global boundary.

ICT promises a fundamental change in all aspects of our lives, including knowledge dissemination, social interaction, economic and business practices, political engagement, media, education, health, leisure and entertainment.

In India ICT applications such as Warana, Dristee, Sari, Sks, E-Chaupal, Cybermohalla, Bhoomi, E-Mitra, Deesha, Star, Setu, Friends, E-Seva, Lokmitra, E-Post, Gramdoot, Dyandoot, Tarahaat, Dhan, Akshaya, Honeybee, Praja are in functioning for rural development.

ICT offers an opportunity to introduce new activities, new services and applications into rural areas or to enhance existing services. With more than 70% of the Indian population living in rural areas and earns its live hood by agriculture and allied means of income.

ICTs can play a significant role in combating rural and urban poverty and fostering sustainable development through creating information rich societies and supporting livelihoods. If ICTs are appropriately deployed and realize the differential needs of urban and rural people, they can become powerful tools of economic, social and political empowerment.

This paper introduces the application of ICT for rural development. The paper aims at improving the delivery of information to rural masses such as: technology information, marketing information, and information advice. This paper focuses digital divide and poverty eradication, good governance and the significance of internet for rural development. The paper concludes that ICTs offer the developing country, the opportunity to look ahead several stages of rural development by the use of internet. Effective use of ICT can demolish geographical boundaries and can bring rural communities closer to global economic systems and be of meaningful help to the underprivileged.

Keywords-ICT,RD,digital divide, ,e-governance,procurement, poverty alleviation,empowement.

I. INTRODUCTION

A. ICT Defined

ICT can be interpreted broadly as "technologies that facilitate communication and the processing and transmission of information by electronic means."

B. ICT and Development

According to Nelson Mandela, "Eliminating the distinction between information rich and information poor countries is critical to eliminating other inequalities between North and South and to improving the quality of life of all humanity." (Mandela, 1998).

C. Rural India

"Just as the whole universe is contained in the Self, so is India contained in the villages"…This has been said by **Mahatma Gandhi**, the father of our nation and the visionary architect of India's rural development programme. The villages epitomize the soul of India. Rural India reflects the very essence of Indian culture and tradition.

II. INTERNET

Internet is another great achievement of ICT. Internet is one of the technologies available for global resources and information haring. Internet has become a common resource of the whole mankind, inquiring and sharing information become easier than ever. With Internet, geographical distance and state borders are eliminated.

The Internet can be a powerful democratizing force, offering greater economic, political and social participation to communities that have traditionally been undeserved and helping developing nations meet pressing needs. It needs to ensure that everyone has a chance to share in the benefits of the Digital Age, information technology.

Internet can be a tool for rural development.

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

TABLE I. WORLD-WIDE INTERNET USERS PENETRATION

Country	Users (In Millions)	Population (In Millions)	Percentage (As on 07/12/09)
India	81.00	1156.89	7.00 %
USA	227.71	307.21	74.10 %
Japan	95.97	127.07	75.52 %
China	360.00	1338.61	26.30 %
UK	46.68	61.11	76.40 %
Germany	54.22	82.32	65.90 %
Canada	25.08	33.48	9.80 %
France	43.10	62.15	69.30 %
Australia	17.03	21.26	61.00 %
Russia	45.25	140.04	32.31 %
Indonesia	30.00	240.27	12.50 %

www.internetworldstats.com/stats.htm

Above table depicts that there is need to promote internet uses in developing country like India.

III. IMPACT OF ICT ON SOCIETY

History has seen a move from the agricultural society, through the industrialized society and to the information society. Society is better informed as a result of developments in ICT. Now a day ICT is used in daily practices. For ex. training, education and entertainment these help society to a great extent. There is impact of increased availability of information using ICT on a variety of public services.

A. Digital Divide

ICTs threaten to expand the already wide socio-economic gap between urban and rural populations in developing countries. Simultaneously offering opportunities to reduce it. ICT implementation can be successfully adapted for the development of challenges faced by rural communities. Such implementations can be evaluated so that the often unexpected and desirable results that emerge can be revealed and accounted for.

B. E-Governance and Empowerment

The poverty can be adequately addressed by effective use of e-governance and ICT application in environmental management. Improved governance by using ICT can have direct impact in reducing poverty and improving the environment.

ICT can contribute fostering empowerment and participation by making government processes more efficient and transparent by encouraging communication and information sharing among rural and marginalized people.

C. ICT and Agriculture

The vast majority of poor people lives in rural areas and derives their livelihoods directly or indirectly from agriculture. ICTs can deliver useful information to farmers about agriculture like crop care and animal husbandry, fertilizer and feedstock inputs, pest control, seed sourcing and market prices.

The ICT application in agriculture sector is still very limited in India, and full range of potential benefits that such technology can provide us yet to be realized.

D. ICT and Poverty Alleviation

ICT applications in developing countries are often part of an overall strategy for economic growth. The role of ICT in poverty reduction is not limited to reducing income poverty, but also includes non-economic dimensions - in particular empowerment.

E. ICT and Health

Health care is one of the most promising areas for poverty alleviation. ICTs are being used in India to facilitate remote consultation, diagnosis and treatment.

Delivering health care with ICTs enables health care professionals and institutions to address the critical medical needs of rural communities, especially those in remote locations and those that lack qualified medical personnel and services.

F. ICT for Education

Moreover, appropriate use of ICTs in the classroom fosters critical, integrative and contextual teaching and learning; develops information literacy (the ability to locate, evaluate and use information). Thus, it improves the overall efficiency of the delivery of education in schools and educational management institutions at the national, state/provincial and community level.

The use of ICTs in education aims to improve the quality of teaching and learning as well as democratize the access to education.

G. ICT for Economic Development

Information and Communication Technology has a vital role in connecting the rural community to outside world for exchange of information, a basic necessity for economic development. Effective use of ICT can demolish geographical boundaries and can bring rural communities closer to global economic systems and be of meaningful help to the underprivileged.

H. Employment Opportunities

Poor people in rural localities have lack of opportunities for employment because they often do not have access to information about them. One use of ICTs is to provide on-line services for job placement through electronic labor exchanges in public employment service or other placement agencies.

IV. CHALLENGES AND ISSUES

The barriers that currently limit the development of rural areas include Distance barriers: to access to administrative and governmental structures, Economic barriers: to access to wider business and labor markets, Social barriers: to information, education facilities, health and social services,

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

Information barriers : many rural areas and their amenities are undiscovered, unknown for the outer world.

For successful implementation of E- rural development it is necessary to define ICT requirements and issues for rural areas like vision, policy, awareness, technology, Infrastructure, services, applications. It is extremely important to focus on all aspects of rural development. For introduction and implementation of E-rural policy it is necessary to establish cooperation of different sectors of government and on national level to introduce coherent E-rural policy, establish formal platform, supporting exchange information and knowledge which will join researchers, developers, regional and local government and which could coordinate research and implementation activity and which can also support E-rural policy □

Information-and-ICT initiatives are political. The effectiveness and potential of ICTD initiatives can be inhibited or circumscribed by national and/or local power relations. Political awareness and analysis is an important aspect of ICT4D planning at all levels.

Harnessing ICTs for human development requires awareness-raising and constituency-building across all levels of society.

The challenge for governments is to ensure the convergence of their initiatives and those taken up by various donors, multilaterals, NGOs and other organizations and to address the digital divide.

V. Discussion / Conclusions

A. *Agriculture*

- The farmers are also not able to know about the prices prevailing in other markets, as the Market Committees are able to disseminate information mostly in respect of their own markets.
- The availability of prompt and reliable market information about what is happening in the market, what quantities are arriving and what prices are quoted for different commodities considerably improves the decision making capability of the farmers and strengthens their bargaining power.

B. *Education*

- With introduction of ICT based education at school level our children and youngsters will grow as "Computer kids". Their exposure will get increased due to which the knowledge level will get definitely improved.
- It enhances the quality of education.

C. *Health Services*

- Delivering health care with ICTs enables health care professionals and institutions to address the critical medical needs of rural communities, especially those in remote locations and those that lack qualified medical personnel and services.
- General knowledge and awareness regarding health can be promoted among the people.

D. *Commerce and Trade*

- Still the benefits are not percolating down to the farmers, as they are unable to plan their strategies for sale of their produce at remunerative prices, in the absence of correct and timely market information and advice about arrivals, prices, market trend, etc.
- In terms of market opportunities emerging agricultural technologies are increasingly information intensive and the rural poor must now cope with increasingly sophisticated input and output markets.

E. *Local Governance and Community life*

- ICT can empower rural communities (particularly marginalized group) and give them "a voice" which permits them to contribute to the development process.
- With ICT government can serve the public better. Computers and the Internet make it possible for government to contact and transact with lots of people at the same time.

VI. Suggestions / Recommendations

- ICT should be harnessed for the benefit of ordinary farmers.
- ICT policy is the most important factor on the introduction of information communication technology to the rural mass with a view to empower them by providing all types of information and communication with regard to their day to day life.
- Awareness building on ICT usage should be conducted to the rural mass through the institutions such as gram panchayat, government office, schools, ICT Centers etc.
- Government has to encourage the software developers to develop software packages in their respective mother tongue.
- While the dedicated and qualified English teachers are appointed to the schools to enhance the English knowledge. Computer knowledge has to be given in their mother tongue with the support of English upto the students are able to understand and handle the computers.
- Established cyber cafes in rural areas to facilitate the rural people who cannot purchase computers by themselves. To encourage the private sector to establish cyber café to provide ICT services in the rural areas.
- It is necessary to open dialogue and professional discussions to create awareness on ICT in the rural areas.

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

- To support and facilitate to develop locally relevant contents for the Internet.
- Significant attention needs to pay for the development of business models for content creation in the rural computing context.

VII. STRATEGIES

- Adopt one village one computer scheme.
- Start Community learning and Information Centers (CLIC) centers.
- Establish Market Information Centers in remote areas.
- Establish Tele Centers in remote areas.
- Encourage the use of computers and Internet in rural areas.
- Establish Information Technology (IT) parks in remote areas.
- Government has to reduce taxation of ICT-related components, products and services.
- Establish partnerships with NGOs engaged in awareness and innovative for ICT4RD.
- Explore the use of Free and Open Source Software (FOSS).
- Explore the use of local language software's.
- Promote the benefits of ICTs to private sector and academic institutions, and encourage computerization.
- Begin basic ICT skills workshops for all rural students at tertiary level.
- Encourage ICT awareness programmes, especially among primary and secondary school students in rural areas.
- Promote ICT-related courses at university/college level and expand the base of supportive certificate and diploma level at college level.
- Connect schools, universities and research centers to national and international distance education facilities, national and international databases, libraries, research laboratories and computing facilities.
- Encourage corporations to appreciate ICT competent staff and conduct/sponsor ICT training for staff members/professionals.
- Encourage assessment and promotion of civil servants to include ICT competency.
- Encourage ICT4D research and development and partnership with the private sector and international educational/research centers.

VIII. THE ROAD AHEAD

Rural Development forms an important agenda of the Government. However, the uptake of e-governance in the Rural Development sector has been relatively slow. The main reasons for this are poor ICT infrastructure in rural areas, poor ICT awareness among agency officials working in rural areas and local language issues. Efforts are, however, on to extend infrastructure up to village level. Already, many states have gone ahead to provide connectivity up to block level. This has helped in taking the e-governance efforts further closer to the people.

The important requirement of establishing infrastructure in rural areas is now being taken up as a high-agenda project after the President of India envisioned the idea of providing urban amenities in rural areas (PURA). PURA (Provision of urban amenities in Rural Areas) has been conceived as a scheme under MoRD and envisages to achieve its objective by bridging the various kinds of divide that exists between rural and urban areas by providing four major kinds of connectivity to rural areas: physical (road, power), electronic (telecommunication, internet), knowledge and market. With the provision of such connectivity, it is hoped that the benefits of e-governance in the Rural Development sector would reach its true beneficiaries.

Crucial success factors to realize this dream are strong political & administrative will, Government Process Reform, capacity building of provider (government functionaries) and consumer (rural citizens), utilization of ICTs as a medium to share information and deliver services that are demand-driven and people-centric.

ACKNOWLEDGMENT *(HEADING 5)*

We are thankful to Hon. Ashok Chavan (Chief Minister, Maharashtra) India, Society members of Shri. Sharada Bhawan Education Society, Nanded. Also thankful to Shri. Jaiprakash Dandegaonkar (Ex-State Minister, Maharashtra), Society members of Bahiri Smarak Vidyalya Education Society, Wapti for encouraging our work and giving us support.

Also thankful to our family members and our students.

REFERENCES

[1] Annual report, 2002-2003, Ministry of Rural Development Government of India.

[2] Data at Nasscom www.nasscom.org

[3] E-Seva, an information brochure of Department of Information Technology and communications 2001.

[4] http://ruralinformatics.nic.in

[5] Governance for sustainable human development UNDP

[6] What is good governance UN-ESCAP

[7] Keith Yeomans ICTs in India, September 2001.

[8] Roger Harris, "ICT for poverty Alleviation Framework", December 2002

[9] 'Information and communications Technology for Development.' A source book for Parliamentarians.

[10] Rural Informatics in India – An approach paper.

[11] Information for Development (i4d) Vol. II No. 12 December 2004.

[12] www.i4donline.net

[13] http://natfm.ac.in/journal/rural.doc

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

[14] Dr Yusaf Samiullah & Mr. Srinivasa Rao, "Role of ICTs in Urban and Rural Poverty Reduction".

[15] Journal of the Eighth National Conference on e-Governance 3-5 February, 2005.

[16] Bastavee Barooah _October 2003, Significance of Internet Access for Rural India.

AUTHORS PROFILE

Dr. N.V.Kalyankar
Principal
Yeshwant Mahavidyalaya, Nanded (Maharashtra)

Completed M.Sc. (Physics) from Dr.B.A.M.U, Aurangabad. In 1980 he joined as a lecturer in department of physics at Yeshwant Mahavidyalaya, Nanded. In 1984 he completed his DHE. He completed his Ph.D. from Dr.B.A.M.U, Aurangabad in 1995. From 2003 he is working as a Principal to till date in Yeshwant Mahavidyalaya, Nanded. He is also research guide for Physics and Computer Science in S.R.T.M.U, Nanded. He is also worked on various bodies in S.R.T.M.U, Nanded. He also published research papers in various international / national journals. He is peer team member of NAAC (National Assessment and Accreditation Council, India). He published a book entitled "DBMS concepts and programming in FoxPro". He also got "Best Principal" award from S.R.T.M.U, Nanded in 2009. He is life member of Indian National Congress, Kolkata (India). He is also honored with "Fellowship of Linnean Society of London (F.L.S.)" on 11 November 2009.

Dr.S.B.THorat
M.E. (Computer Science & Engg.)
M.Sc. (ECN), AMIE, LM-ISTE, Ph.D. (Comp.Sci. & Engg.)

He is having 24 years teaching experience. From 2001 he is working as a Director, at ITM. He is Dean of faculty of Computer studies at Swami Ramanand Teerth Marathwada University, Nanded (Maharashtra). Recently he is completed his Ph.D. He attended many national and International conferences. He is having 7 international publications. His interested area are AI, Neural network, Data mining, Fuzzy systems, Image processing.

S.K.Nayak
M.Sc. (Computer Science), D.B.M, B.Ed.

He completed M.Sc. (Computer Science) from S.R.T.M.U, Nanded. In 2000 he joined as lecturer in Computer Science at Bahirji Smarak Mahavidyalaya, Basmathnagar. From 2002 he is acting as a Head of Computer Science department. He is doing Ph.D. He attended many national and international conferences, workshops and seminars. He is having 2 international publications. His interested areas are ICT, Rural development, Bioinformatics.

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

A proof Procedure for Testing Membership in Regular Expressions

Keehang Kwon and Hong Pyo Ha
Dong-A University Department of Computer Engineering
Busan, Republic of Korea

Jiseung Kim
Kyung-IL University Department of Industrial Engineering
Daegu, Republic of Korea

Abstract

We propose an algorithm that tests membership for regular expressions and show that the algorithm is correct. This algorithm is written in the style of a sequent proof system. The advantage of this algorithm over traditional ones is that the complex conversion process from regular expressions to finite automata is not needed. As a consequence, our algorithm is simple and extends easily to various extensions to regular expressions such as timed regular expressions or regular languages with the intersection.

Key words: *regular expressions, proof theory, linear logic, algorithm*.

1. Introduction

Since its introduction, regular expressions [2,5] have gained much interest for applications such as text search or compiler components. One important question is, given a string *w* and a regular expression *r*, to decide whether *w* is in the set denoted by *r*. Testing membership in a regular expression has traditionally concentrated on converting a regular expression to finite automata. Such a conversion technique is unsatisfactory for at least two reasons. First, the conversion process itself requires a lot of extra overloads. Second, the conversion technique has only a limited number of applications and does not extend well to various – even simple – extensions (regular expressions with time [1], regular expressions with intersection, etc) to regular expressions.

Little has been studied about the algorithms for testing membership for regular expressions themselves. This paper introduces such an algorithm. It is simple, easy to understand, nondeterministic and has some resemblance to the proof theory of intuitionistic linear logic[3] . In addition, it is a simple matter to observe that this algorithm extends well to other extensions to regular expressions.

In this paper we present our algorithm, show some examples of its working, and discuss further improvements. The remainder of this paper is structured as follows. We describe our algorithm in the next section.

In Section 3, we present some examples. Section 4 concludes the paper with some considerations for further improvements.

2. The language

The regular expression is described by *r*-formulas given by the syntax rules below:

$$r ::= \emptyset \mid \varepsilon \mid a \mid r \cdot r \mid r + r \mid r^*$$

In the rules above, an alphabet a represents the set $\{a\}$. ε represents $\{\varepsilon\}$. $\emptyset$ represents the empty set. $r \cdot s$ represents the concatenation of two sets r and s. $r + s$ represents the union of r and s. The Kleene closure of r - r^* - indicates there are any number of r.

We often write rr in place of $r \cdot r$. The degree $d(r)$ of a regular expression r is defined as follows: $d(\emptyset)=0$, $d(a)=1$, $d(r\cdot s) = d(r + s) = d(r) + d(s) +1$,$d(r^*) = d(r) +1$. Further, $d(r_1,\ldots,r_n) = d(r_1)+\ldots+d(r_n)$ where $r_1 ,\ldots,r_n$ is a sequence of regular expressions.

The question of whether a string is a member of a regular expression is quite interesting. Such an algorithm needs to cope with the following: (1) the associativity of the operators, e.g., $abc \in a \cdot (b \cdot c)$, $abc \in (a \cdot b) \cdot c$, and (2) the multiplicity of the Kleene closure, *e.g.*, $aaa \in a^*\cdot a^*$, in an elegant fashion. We will present an algorithm for this task in the style of a proof system.

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

Let w be a regular expression and $r_1, \ldots, r_n$ be a list of regular expression. Then a *sequent* of the form $r_1, \ldots, r_n \vdash w$ – the notion that is an element of the concatenations of $r_1, \ldots, r_n$ – is defined constructively by two axioms and eight inference rules. This is shown below.

Algorithm for Testing Membership

$$\frac{}{a \vdash a}\ Axiom1 \qquad \frac{}{\vdash \epsilon}\ Axiom2$$

$$\frac{\rho, \Psi, \Delta \vdash w}{\rho \cdot \Psi, \Delta \vdash w} \cdot L \qquad \frac{\Delta_1 \vdash w_1}{\Delta_1, \Delta_2 \vdash w_1 w_2} \cdot R$$

$$\frac{\Delta \vdash w}{\rho, \Psi \vdash w} \epsilon L \qquad \frac{\Delta \vdash w}{\rho^*, \Delta \vdash w} WL$$

$$\frac{\rho^*, \rho^*, \Delta \vdash w}{\epsilon, \Delta \vdash w} CL \qquad \frac{\rho, \Delta \vdash w}{\rho^*, \Delta \vdash w} DL$$

$$\frac{\rho, \Delta \vdash w}{\rho + \Psi, \Delta \vdash w} + L_1 \qquad \frac{\Psi, \Delta \vdash w}{\rho + \Psi, \Delta \vdash w} + L_2$$

In the above rule, a is an alphabet, Δ denote a list of regular expressions and ρ, ψ denote a single regular expression. An inference rule can be read as follows: if all the sequents above are true, then the sequent below is true. A sequent $\Delta \vdash w$ has a *proof* if $\Delta \vdash w$ can be obtained from the axioms by applying the inference rules. In dealing with ρ^* construct, the proof system can either discard it, use ρ once, or use ρ at least twice.

Let us refer to the above collection of axioms and inference rules as *DS*. The following theorem shows the sound and completeness of the proof system *DS*.

Theorem 2.1: Let $r_1, \ldots, r_n$ be a list of regular expressions and let w be a string. Then, w is an element of $r_1 \ldots r_n$ if and only if $r_1, \ldots, r_n \vdash w$ has a proof in *DS*.

Proof. The reverse direction is straightforward. In the forward direction, we prove the theorem by an induction on the degree of the sequence $r_1 \ldots, r_n, w$. If the degree is 1 or 2, then it must be of the form $\varepsilon \in \varepsilon$, $\varepsilon \in$,or $a \in a$ where a is an alphabet. It is easy to see that the theorem is true.

If the degree is greater than 2, we consider the cases for the structure of r_1.

If r_1 is a, then it must be the case that w is of the form aw' and $w' \in r_2, \ldots, r_n$ where w' is a (possibly empty) string. By the hypothesis, $r_2 \ldots r_n \vdash w'$ has a proof. Putting the proofs for $a \vdash a$ and $r_2 \ldots r_n \vdash w'$ using a $\cdot$ R rule, we obtain a proof satisfying the theorem.

If r_1 is $r+s$, then it must be the case that $w \in r, r_2 \ldots r_n$ or $w \in s, r_2 \ldots, r_n$. Consider the former case. By the hypothesis, $r, r_2, \ldots r_n \vdash w$ has a proof. Putting this proof using a $+L_1$ rule, we obtain a proof satisfying the theorem. The same argument can be supplied for the latter case.

The arguments for rs follow a similar pattern. For the case that r_1 is of the form r^*, we have to consider the three cases depending on whether r is never used, used once, or used more than once. Consider the case where r is never used. Then w must be an element of $r_2, \ldots, r_n$. By the hypothesis, $r_2, \ldots, r_n \vdash w$ has a proof. Putting this proof using a WL rule, we obtain a proof satisfying the theorem. This is shown below.

$$\frac{r_2, \cdots, r_n \rightarrow w}{r^*, r_2, \cdots, r_n \rightarrow w} WL$$

If r is used once, then there must be a string z such that $w = zw'$, $z \in r$, and $w' \in (r_2, \ldots, r_n)$. By the hypothesis, both $r \vdash z$ and $r_2, \ldots, r_n \vdash w'$ have proofs. Putting the proofs for $r \vdash z$ and $r_2, \ldots, r_n \vdash w'$ using the rules, we obtain a proof satisfying the theorem.

This is shown below.

$$\frac{\dfrac{r \rightarrow z \qquad r_2, \cdots, r_n \rightarrow w'}{r, r_2, \cdots, r_n \rightarrow zw'} \cdot R}{r^*, r_1, \cdots, r_n \rightarrow zw'} DL$$

The arguments for the remaining case follow a similar pattern. The additional observation is that the *CL* rule is needed for this case.

3. Examples

This section describes the use of our algorithm. An example is provided by the following proof of $aa \in a^*$.

It is interesting to note that the *CL* rule is used to control the multiplicity of a^*.

$$\frac{\dfrac{\dfrac{a \rightarrow a}{a^* \rightarrow a} DL \qquad \dfrac{a \rightarrow a}{a^* \rightarrow a} DL}{a^*, a^* \rightarrow aa} \cdot R}{a^* \rightarrow aa} CL$$

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

Another example of our algorithm is provided by the following proof of the sequent $ca \in (b+c)\cdot a$

$$\dfrac{\dfrac{\dfrac{c \to c}{b+c \to c}+L_2 \quad a \to a}{b+c, a \to ca}\cdot R}{(b+c)\cdot a \to ca}\cdot L$$

A computation process typically searches for a proof from the bottom-up in a sequent calculus for reasons of efficiency. Thus, given a conclusion sequent, it attempts to find its proof from bottom-up.

4. Conclusion

We have described an algorithm for testing membership in regular expressions. The advantage of this algorithm is that it does not require the complex conversion process to finite automata. As a consequence, it extends easily to various extensions to regular expressions. For example, our algorithm extends easily to the one that deals with algebraic laws, *i.e.*, regular expressions with variables [5]. Two regular expressions with variables are equivalent if whatever expressions we substitute for the variables, the results are equivalent. For example, $\forall L \forall M(L+M=M+L)$.

Regarding the performance of our algorithm, non-determinism is present in several places of this algorithm. In particular, there is a choice concerning which way the text is split in the *·R rule*. Hodas and Miller [4] dealt with this rule by using IO-model in which each goal is associated with its input resource and output resource. The idea used here is to delay this choice of splitting as much as possible. This observation leads to a more viable implementation. Our ultimate interest is in a procedure for carrying out computations of the kind described above. It is hoped that these techniques may lead to better algorithms.

5. Acknowledgements

This paper was supported by Dong-A University Research Fund.

References

[1] E. Asarin and P. Caspi, and O. Maler, "A Kleene theorem for timed automata", Logic in Computer Science, 1997, pp.160-171.

[2] S.C. Kleene, Introduction to Metamathematics, North-Holland, Amsterdam, 1964.

[3] J.Y. Girard, "Linear logic", Theoretical Computer Science, vol.50, pp.1-102, 1987.

[4] J.Hodas and D. Miller, " Logic programming in a fragment of intuitionistic linear logic", Journal of Information and Computation, 1992. Invited to a special issue of submission to the 1991 LICS conference.

[5] J.E.Hopcroft, R.Motwani, and J.D. Ullman, Automata Theory, Languages and Computation, Ad.

Impact of Random Loss on TCP Performance in Mobile Ad-hoc Networks (IEEE 802.11): A Simulation-Based Analysis

Shamimul Qamar
Department of Computer Sceience
CAS, King Saud University
Riyadh,Saudi Arabia

Kumar Manoj
ECED,DPT
I.I.T Roorkee
India

Abstract-Initially TCP was designed with the notion in mind that wired networks are generally reliable and any segment loss in a transmission is due to congestion in the network rather than an unreliable medium (The assumptions is that the packet loss caused by damage is much less than 1%) . This notion doesn't hold in wireless parts of the network. Wireless links are highly unreliable and they lose segments all the time due to a number of factors. Very few papers are available which uses TCP for MANET. In this paper, an attempt have been made to justify the use of TCP variants (Tahoe and Reno) for loss of packet due to random noise introduces in the MANET. For the present analysis the simulation has been carried out for TCP variants (Tahoe and Reno) by introduces 0%, 10%, 20% and 30% noise.The comparison of TCP variants is made by running simulation for 0%, 10%, 20% and 30% of data packet loss due to noise in the transmission link and the effect of throughput and congestion window has been examined. During the simulation we have observed that throughput has been decreased when a drop of multiple segments happens, further we have observed in the case of TCP variant (Reno) throughput is better at 1% (Figure 5) which implies a network with short burst of error and low BER, causing only one segment to be lost. When multiple segments are lost due to error prone nature of link, Tahoe perform better than Reno (Figure 13), that gives a significant saving of time (64.28%) in comparison with Reno (Table 4). Several simulations have been run with ns-2 simulator in order to acquire a better understanding of these TCP variants and the way they perform their function. We conclude with a discussion of whether these TCP versions can be used in Mobile Ad hoc Network?

Index Term- TCP, Tahoe, Reno, MANET, ns-2, Random noise

1. Introduction

TCP is the widely used for Internet traffic. In the wired network losses are mostly due to congestion. In practice, losses may also be caused from noisy links. This is especially true in case of radio links, e.g. in cellular network, Mobile Ad-hoc network or in satellite links. A link may become in fact completely disconnected for some period of time or it may suffer from occasional interferences (due to shadowing, fading etc.) that cause packets to contain errors and then to be dropped resulting low throughput. Transport protocols should be independent of the technology of the underlying network layers. It should not care whether IP is running over fiber or radio. But in reality, it does matter, since most TCP implementations that have been designed for wired networks, don't cope very well on wireless networks. Different version of TCP in a local network with lossy link are compared and analyzed in [7].Several scheme have been reported in the literature, to alleviate the effect of losses on TCP performances over wireless network (or networks with lossy link) [8]. Very few papers are available which uses TCP for MANET This paper deals with for Tahoe and Reno TCP variants and its variation of error by introducing different level of noise . The ns-2 network simulator used to understand the behaviors and characteristics of TCP variants. Wireless links are highly unreliable and they lose segments all the time due to a number of factors. These include fading, interference, higher bit error rate and mobility related processes such as handovers.

In this paper, we will make a comparison between Tahoe and Reno TCP variants in a noisy environment. Our simulation show that Tahoe copes well in a high segment loss environment. Two algorithms, Slow Start and Congestion Avoidance, modify the performance of TCP's sliding window to solve some problems relating to congestion in the network [4]. The idea of congestion control is for the sender to determine the capacity that is available on the network. The sender, in standard TCP implementation, keeps two state variables for congestion control: a slow-start/congestion window, *cwn*d, and a threshold size, *ssthres*h. These variables are used to switch between the slow start and congestion avoidance algorithms. Slow start provides a way to control initial data flow at the beginning of a communication session and during an error recovery. This is based on received acknowledgements. Congestion avoidance increases the *cwnd* additively, so that it grows by one segment for each round trip time. These two algorithms are implemented together and work as if they were one. The combined

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

effect of these two algorithms on the *congestion window* is shown in Figure 1. In this paper we report the simulation results of different scenarios. In Section-2**,** a summary of Tahoe and Reno TCP transport protocols has been reported.

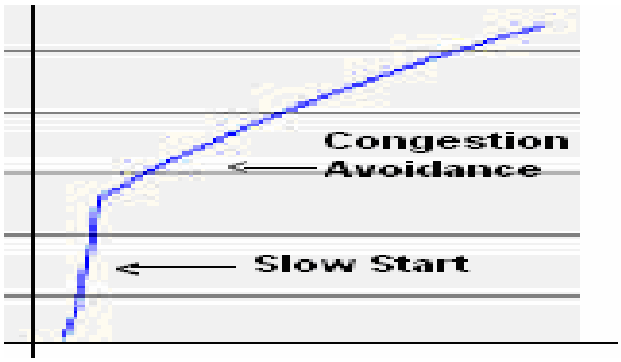


Figure 1. Congestion window

The simulation software and the network simulation setup are described in section 3. The comparison of simulation results is presented in section 4. The conclusion has been reported in section 5.

2. TCP Transport Protocol

2.1 Tahoe TCP

Modern TCP implementations contain a number of algorithms aimed at controlling network congestion while maintaining good user throughput. Early TCP implementations followed a go-back- model using cumulative positive acknowledgment and requiring a retransmit timer expiration to re-send data lost during transport. These TCPs did little to minimize network congestion. The Tahoe TCP implementation added a number of new algorithms and refinements to earlier implementations. The new algorithms include *Slow-Start*, *Congestion Avoidance*, and *Fast Retransmit* [3]. The refinements include a modification to the round-trip time estimator used to set retransmission timeout values. All modifications have been described elsewhere. The fast retransmit algorithm is of special interest in this paper because it is modified in subsequent versions of TCP. With Fast Retransmit, after receiving a small number of duplicate acknowledgments for the same TCP segment (*dup ACKs*), the data sender infers that a packet has been lost and retransmits the packet without waiting for a retransmission timer to expire, leading to higher channel utilization and connection throughput.

2.2 Reno TCP The Reno TCP implementation retained the enhancements incorporated into Tahoe, but modified the Fast Retransmit operation to include *Fast Recovery* [3]. The new algorithm prevents the communication path (“pipe”) from going empty after Fast Retransmit, thereby avoiding the need to slow-start to re-fill it after a single packet loss. Fast recovery operates by assuming each dup ACK received represents a single packet having left the pipe. Thus, during Fast recovery the TCP sender is able to make intelligent estimates of the amount of outstanding data. Fast Recovery is entered by a TCP sender after receiving an initial threshold of dup ACKs. This threshold, usually known as *tcprexmtthresh*, is generally set to three. Once the threshold of dup ACKs is received, the sender retransmits one packet and reduces its congestion window by one half. Instead of slow-starting, as is performed by a Tahoe TCP sender, the Reno sender uses additional incoming dup ACKs to clock subsequent outgoing packets. In Reno, the sender' s *usable* window becomes where is the receiver' s advertised window, is the sender' s congestion window, and is maintained at until the number of dup ACKs reaches *tcprexmtthresh*, and thereafter tracks the number of duplicate ACKs. Thus, during Fast Recovery the sender “inflates” its window by the number of dup ACKs it has received, according to the observation that each dup ACK indicates some packet has been removed from the network and is now cached at the receiver. After entering Fast Recovery and retransmitting a single packet, the sender effectively waits until half a window of dup ACKs have been received, and then sends a new packet for each additional dup ACK that is received. Upon receipt of an ACK for new data (called a “recovery ACK”), the sender exits Fast Recovery by setting to Reno's fast recovery algorithm is optimized for the case when a single packet is dropped from a window of data. The Reno sender retransmits at most one dropped packet per round-trip time as shown in Table[1]. Reno significantly improves upon the behavior of Tahoe TCP when a single packet is dropped from a window of data, but can suffer from performance problems when multiple packets are dropped from a window of data. This is illustrated in the Table[2] with three or more dropped packets as shown in Table[3].

Table 1: Round trip times in seconds (at sink nodes)

Minimal RTT(CN,ON,SPID)	0.303733(0,4,2)
Maximal RTT(CN,ON,SPID)	0.594773(0,4,49)
Average RTT	0.5565596345

Table 2.Average generated & received packet of Tahoe & Reno for different niose.

Tahoe (at source node)				
	0%	10%	20%	30%
Generated packets	290	63	61	21
Average Packet Size	538.275	561.0084	606.0377	628.2353
Reno (at sink node)				
	0%	10%	20%	30%
Received Packets	290	56	45	13

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

Average Packet Size	538.275	531.0714	606.0377	501.5385

The problem is easily constructed in our simulator when a Reno TCP connection with a large congestion window suffers a burst of packet losses after slow-starting in a network with drop-tail gateways (or other gateways that fail to monitor the average queue size).

3. Simulation Setup

This section describes four simulation scenarios for 0%, 10%, 20% and 30% of data packet loss due to noise in the transmission link. Each set of scenarios is run for Tahoe, Reno, New-Reno and SACK TCP. The results in this paper are based on simulations using the ns-2 network simulator. The network is setup as per the network diagram in the Figure 2 such as there are 5 nodes from n0 to n5 and node n0 and n2 are connected by bidirectional link of 2 Mbps bandwidth and a propagation delay of 10ms. We have implemented a DropTail queue between node n2 and n3. FTP traffic is sent from node n0 to n4 via shared link (n2-n3) by using TCP transport protocol and CBR traffic is sent from node n1 to n5 using same shared link (n2-n3) by using UDP protocol.

The various simulation parameters used for the present analysis for both (Tahoe and Reno) are like no. of nodes (6), simulation length in seconds(10.47), no. of sending nodes (3), no. of receiving nodes(3), no. of generated packet(593), no. of forwarded packet(1189), number of forwarded bytes(325200), average packet size(548.398).

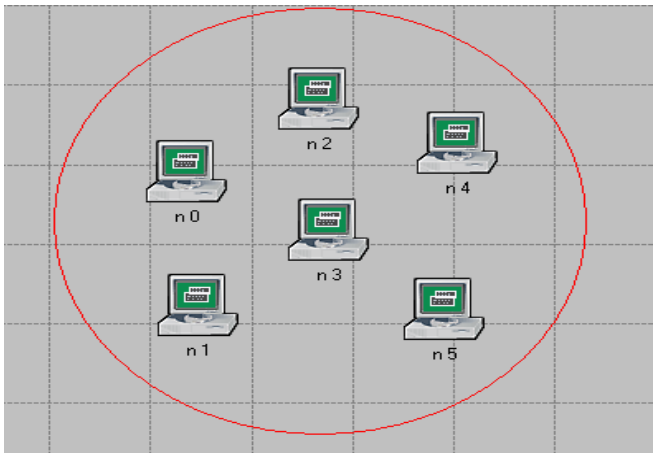


Figure 2. Proposed system network

The simulation has been carried out for 141 simulation seconds in all four scenario viz; (0%, 10%, 20% and 30%) for Tahoe and Reno TCP varients. In the first scenario the network is simulated without noise at the shared link whereas in the second, third and fourth scenario we have introduced a noise of 10%, 20% and 30% are introduced in the shared link (forward link of n2 to n3) respectively.

4. Analytical and Simulation Comparisons

In this section we have presented the simulation results for the scenarios mentioned above.

On comparing the throughput of TCP Reno and Tahoe at without loss (0% noise) the throughput remains same as shown in Figure 3-4, whereas on increasing the noise (1%) the throughput of Reno is better as compared with the throughput of Tahoe as shown in Figure 5. Fast Retransmit and Fast Recovery in Reno seem to work well when only one segment is lost. But, as we have seen in the simulation

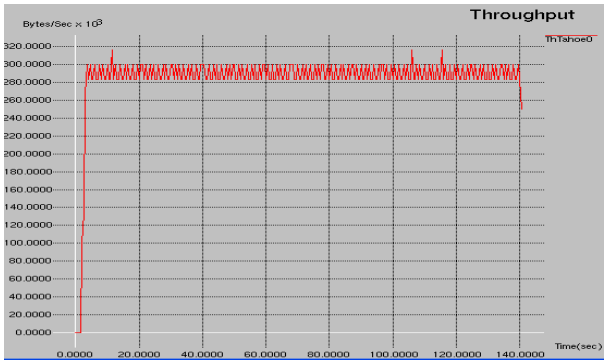


Figure 3. Throughput of TCP Tahoe without loss (0 % noise)

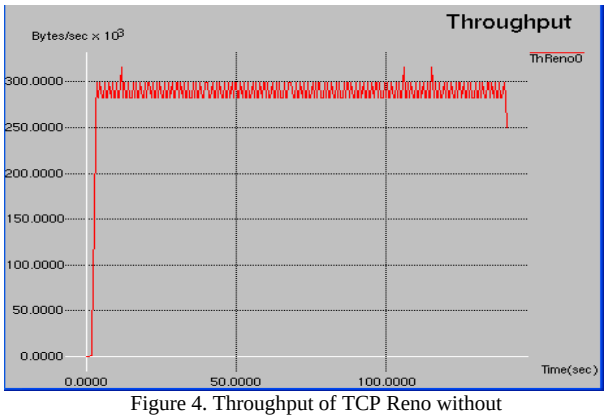


Figure 4. Throughput of TCP Reno without loss (0 % noise)

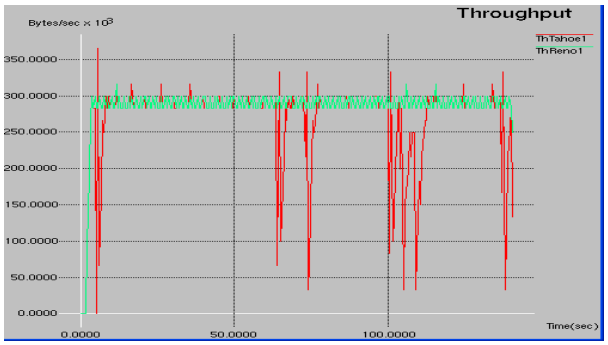


Figure 5. Throughput of TCP Tahoe & Reno with loss (1 % noise)

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

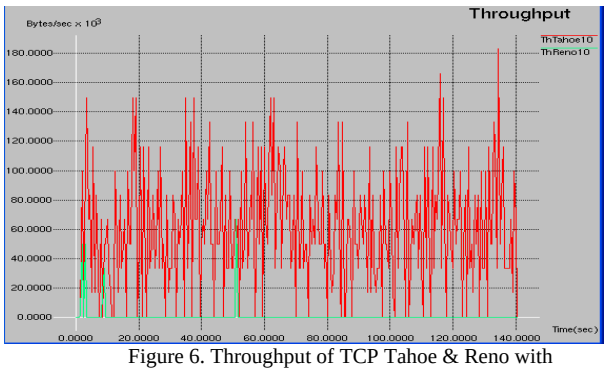


Figure 6. Throughput of TCP Tahoe & Reno with loss (10 % noise)

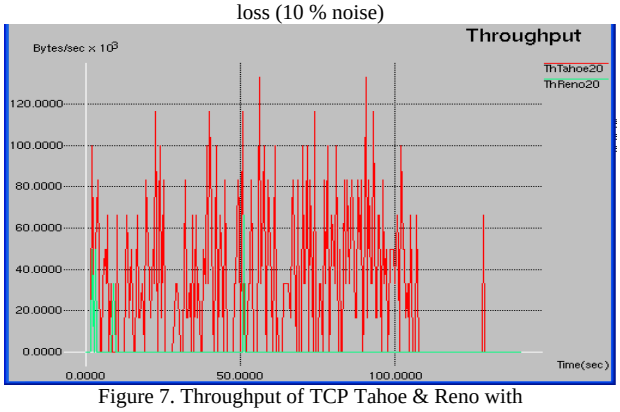


Figure 7. Throughput of TCP Tahoe & Reno with loss (20 % noise)

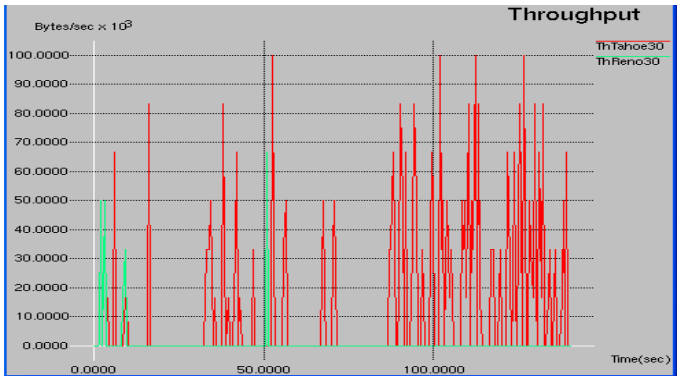


Figure 8. Throughput of TCP Tahoe & Reno with loss (30 % noise)

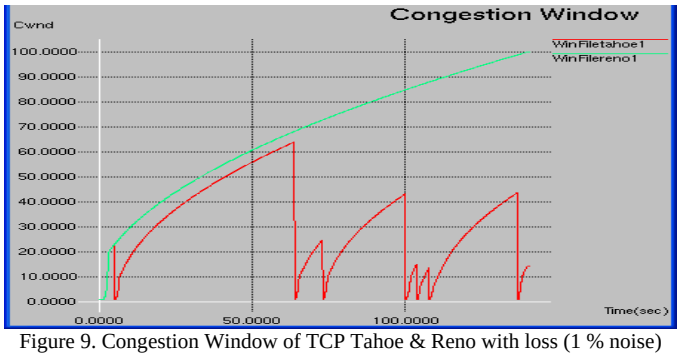


Figure 9. Congestion Window of TCP Tahoe & Reno with loss (1 % noise)

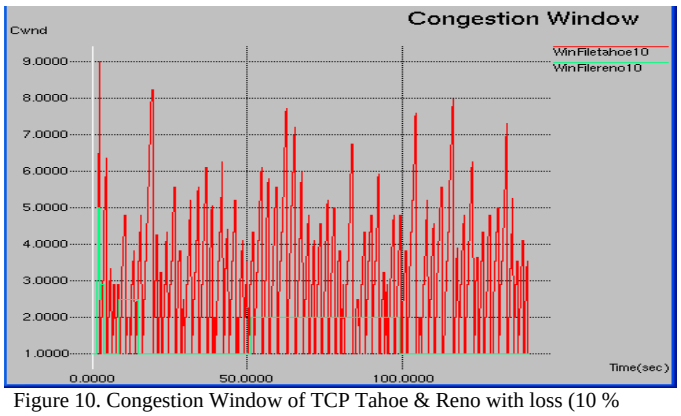


Figure 10. Congestion Window of TCP Tahoe & Reno with loss (10 % noise)

results, some problems arise when a drop of multiple segments happens.

As shown in Figures 10-12, at time 4, 15 and 98 sim-secs when an error causing a loss of multiple segments, a problem arises with Reno TCP's congestion window which results in zero throughputs as shown in Figure 6-8. A loss of multiple segments makes the window close to half of its size for every dropped segment. Besides, the usable window is decreased during the fast recovery phase, as more information is sent; every time a duplicated acknowledgement is received. When the congestion window is divided by two for the second time, the usable window closes to zero, thus blocking the communication and forcing the retransmission timeout. Tahoe TCP does not apply Fast Recovery in order to avoid the problems described previously.

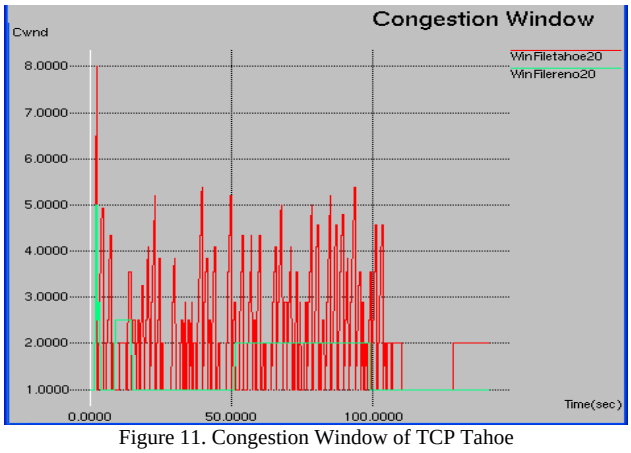


Figure 11. Congestion Window of TCP Tahoe & Reno with loss (20 % noise)

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

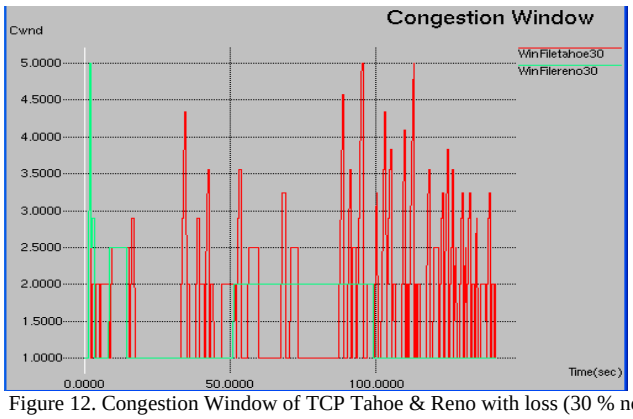


Figure 12. Congestion Window of TCP Tahoe & Reno with loss (30 % noise)

Table 3. Average packet generated and received of Tahoe & Reno variants

Tahoe (at source node)				
	0%	10%	20%	30%
Generated packets	290	76	58	27
Average Packet Size	538.275	564.4755	598.8235	630.9091
Tahoe (at sink node)				
	0%	10%	20%	30%
Received Packets	290	67	44	17
Average Packet Size	538.275	532.5373	528.6364	510.5882

Table 4 End to End delay of Tahoe & Reno variant.

Tahoe (end2end delay in sec)	
Minimal (node,PID)	0 (2,0)
Maximal (node,PID)	0.244373 (2,49)
Average delay	0.05232697133
Reno (end2end delay in sec)	
Minimal delay (CN,ON,PID)	0.151866(4,0,72)
Maximal delay(CN,ON,PID)	0.254(0,4,23)
Average delay	0.183139041

When this variant is used, Fast Retransmit is the only algorithm applied. This means that slow start and congestion avoidance will be working when recovering from an error. Tahoe TCP closes the congestion window when the first error is detected. This stops the communication suddenly, but allows the congestion window, and consequently the usable window, to open exponentially as shown in Figure 9. Then the usable window allows the sender to retransmit the lost segments. So, although the usable window is closed, there is always a mechanism that opens it again. The communication is never blocked as in Reno and no waiting for the retransmission timer is needed.

As we discussed above, Tahoe with little modification can be considered in the scenarios where multiple segments are dropped such as in Mobile Ad hoc Networks.

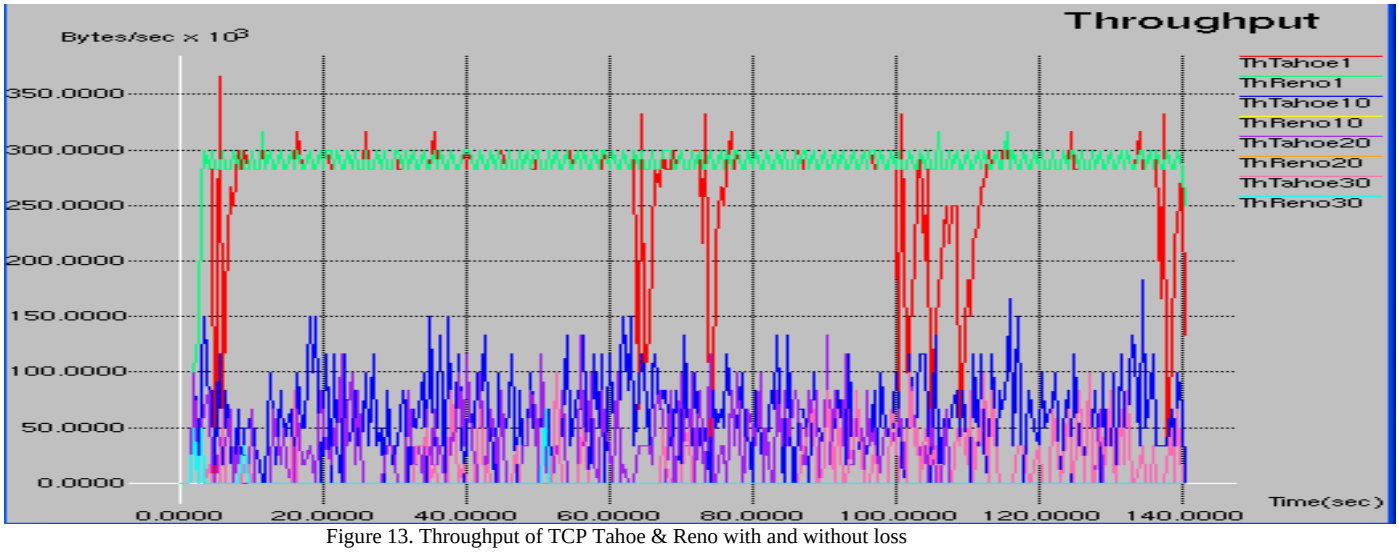


Figure 13. Throughput of TCP Tahoe & Reno with and without loss

5. Conclusion

This paper summarizes a study on TCP variants and its congestion control mechanism. We have analyzed from the simulated results in that Tahoe performs better than Reno in multiple data packets loss that gives a significant saving of time (64.28%) in comparison with Reno as shown in (Table 4). The comparison of TCP variants is made by running simulation for 0%, 10%, 20% and 30% of data packet loss due to noise in the transmission link and the effect of throughput and congestion window has been examined in Figure 13.

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

Thus, TCP Reno implementation could be useful when TCP is working in a reliable network with short bursts of errors and low BER, causing only one segment to be lost. Tahoe can be considered as transport protocol for Mobile Ad hoc network where multiple segments are dropped due to error prone nature of radio link Table [2-3]. Our future work is to focus on the implementation of modified Tahoe in MANET.

6. References

[1] http:// www.isi.edu/nsnam/ns/

[2] TCP Slow Start, Congestion Avoidance, Fast Retransmit and Fast Recovery Algorithms. Network Working Group, RFC: 2001, W. Stevens, January 1997.

[3] Simulation Based Comparisons of Tahoe, Reno and SACK TCP". K. Fall, S. Floyd, 1995.

[4] Computer Networks, The Transport Layer, Andrew S. Tanenbaum, 3rd Edition.

[5] Modified Tahoe TCP for Wireless Networks Using OPNET Simulator", M.N. Akhtar, M.A.O. Barry and H.S.Al-Raweshidy whitepaper NDNet.co.uk. 17 Jan. 2008.

[6] R.Zorzi, A. Chockalingam, R.R. Rao. Throughput analysis of TCP on channels with memory. IEEE J-SAC, 18(7): 1289-1300, 2000.

[7] A. Kuma. Comparative performances analysis of version of TCP in local network with lossy link. Proc. IEEE/ACM Trans. On networking, 6(4): 485-498, 1998

[8] A.C.F. Chan (UST@HK), D.H.K. Tsang, S. Gupta. Performance analysis of TCP in the Presence of random losses/errors. Proc. IEEE ,9th PIMRC'98 825-830, 1998.

Authors:

Dr. Shamimul Qamar, Professor, Computer Science and Information System department, CAS, King Saud University, Riyadh, has been recognized as an eminent scholar in the field of Compuer Engineering. He has done his B.Tech from MMMEC Gorakhpur, M.Tech from AMU, Aligarh and earned his Ph.D. degree from IIT Roorkee, India with highly honorable grade. Prof. Qamar has a wide teaching experience in various Engineering colleges. He has research interests in Image processing, Internet Applications, Multimedia systems, computer network and software engineering. He has published several research papers in reputed national/international Journals and conference. He is also a technical programme committee member in international mobility conference, Singapore. He is a life time member of international association of Engineers and a life member of Indian Society of technical educational. His technical depth and interest resulted in setting up a research lab according to latest technical innovations.

Kumar Manoj (kumardpt@iitr.ernet.in), member of IEEE, ACEEE, IAENG, ISOC, NSBE (USA) received B.Sc. (Engg.), M.Tech. (Electronics). He has published over Fifty nine research papers in national and International journals/conferences and supervised more than 40 projects/ dissertation of M.Tech. & B.Tech. students. He started his career as R & D Engineer in various MNC companies in the field of Power Electronics then joined teaching profession as a Asstt. Prof. in Graphic Era University. He is a visiting faculty of various Govt. Engg. College. His many research papers have been awarded by National and International Committees/Conference. Presently he is pursuing research work at IIT Roorkee, India in the field of Wireless Communication under Ministry of HRD, Government of India fellowship. His research interests include the design and control of personal communication networks, mobile multicasting, protocol design and implementation for a mobile integrated services wireless radio network, and high-speed networking.

Automatic diagnosis of retinal diseases from color retinal images

D.Jayanthi
PG Scholar
Dept of CSE
Sri Venkateswara college of Engineering

N.Devi
Senior Lecturer
Dept of IT
Sri Venkateswara college of Engineering

S.SwarnaParvathi
Senior Lecturer
Dept of IT
Sri Venkateswara college of Engineering

***Abstract*-Teleophthalmology holds a great potential to improve the quality, access, and affordability in health care. For patients, it can reduce the need for travel and provide the access to a super-specialist. Ophthalmology lends itself easily to telemedicine as it is a largely image based diagnosis. The main goal of the proposed system is to diagnose the type of disease in the retina and to automatically detect and segment retinal diseases without human supervision or interaction. The proposed system will diagnose the disease present in the retina using a neural network based classifier.The extent of the disease spread in the retina can be identified by extracting the textural features of the retina. This system will diagnose the following type of diseases: Diabetic Retinopathy and Drusen.**

Keywords:Drusen,Diabetic Retinopathy, retinal diseases, Teleopthamology

I. INTRODUCTION

The World Health Organization estimates that 135 million people have diabetes mellitus worldwide and that the number of people with diabetes will increase to 300 million by the year 2025.Early detection and treatment of these diseases are crucial to avoid preventable vision loss.

Diabetes mellitus (DM) is a chronic, systemic, life-threatening disease characterized by disordered metabolism and abnormally high blood sugar (hyperglycemia) resulting from low levels of the hormone insulin with or without abnormal resistance to insulin's effects. Through computer simulations it is possible to demonstrate that prevention and treatment are relatively inexpensive compared to the healthcare and rehabilitation costs incurred by visual loss or blindness.

Assessment of the risk for development of age-related macular degeneration (ARMD) requires reliable detection and quantitative mapping of retinal abnormalities that are considered as precursors of the disease. Typical signs for the latter are the so-called drusen that appear as abnormal white-yellow deposits on the retina. Color retinal images are used presently to visually identify the presence of drusens. Segmentation of these features using conventional image analysis methods is quite complicated mainly due to the non-uniform illumination and the variability of the pigmentation of the background tissue. Automated detection and analysis can provide vital information about the quantity and quality of the drusens.

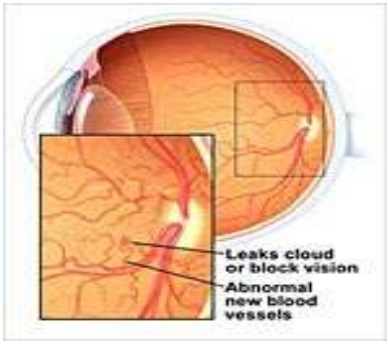


Fig.1 Diabetic Retinopathy

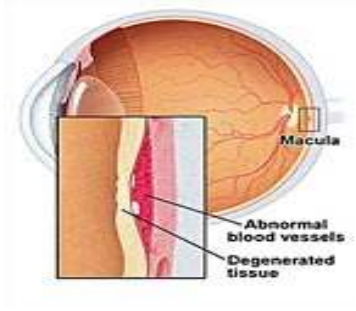


Fig.2 Age-related Macular Degeneration

II. PROPOSED SYSTEM

In the proposed system, a combined method for classifying the type of retinal disease and automatic diagnosis can be done using a neural network classifying technique. The main goal of the proposed system is to diagnose the age-related macular degeneration

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

(drusen) and diabetic retinopathy diseases in the retina and to automatically detect and segment the above diseases without human supervision or interaction. Texture analysis is used to extract the features of the retina. After feature extraction process, a neural network based classifier is used to classify the type of retinal disease and can automatically diagnose the type of disease. The location of drusen and diabetes's can be identified and by extracting the textural features of the retina, the extent of disease spread can be determined.

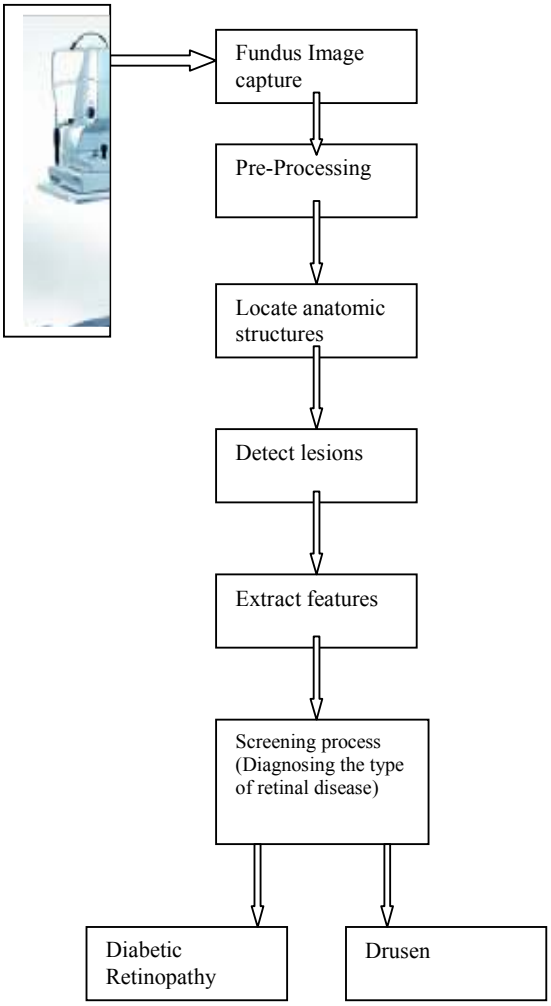


Fig. 3 Block diagram of the proposed system

III. MODULES

In the proposed system, there are four modules, they are as follows:

1.Pre-Processing of a color retinal image.
2.Locating anatomic structures and detecting lesions.
3.Feature Extraction.
4.Classification of retinal disease using Artificial Neural Networks.

A. Pre-Processing

Pre-Processing of retinal image is the first step in the automatic diagnosis of retinal diseases. The problem with retinal image is that the quality of the acquired images is usually not good. So, it is necessary to improve the quality of retinal image. The purpose of pre-processing is to remove the noisy area from retinal image. This is required for the reliable extraction of features and abnormalities as feature extraction and abnormality detection algorithms give poor results in the presence of noisy background..

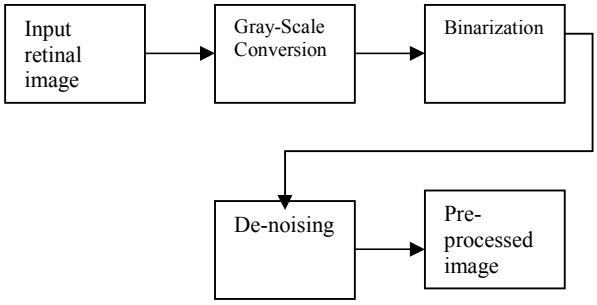


Fig. 4 Block diagram for preprocessing

Steps for pre-processing

1.Divide the input retinal image into non-overlapping blocks.
2.Extract RGB components from the original color retinal image.
3.After gray-level conversion, use histogram equalization to enhance the contrast and to improve the quality of retinal image.
4.Use a large median filter to remove the noise from the image.

B. Locating anatomic structures and detecting lesions

The purpose of locating anatomic structures is to detect the optic nerve based on segmentation of the vascular arcades. Detection of the anatomic structures is fundamental to the subsequent characterization of the normal or disease state that may exists in the retina. The algorithm is based on mathematical morphology and curvature evaluation for the detection of vessel like patterns in a noisy environment. Vessel detection is based on the computation of parameters related to blood flow. In order to define the vessel like pattern, segmentation will be performed with respect to a precise model. In order to differentiate vessels from analogous patterns, a cross curvature evaluation is performed. Vessel like patterns

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

are bright features defined by morphological properties like linearity, connectivity, width and by a specific Gaussian like profile whose curvature varies smoothly along the vessel.

The detection algorithm is based on four steps.
1.Noise Reduction.
2.Linear pattern with Gaussian-like profile improvement.
3.Cross curvature evaluation.
4.Linear filtering.

Our goal is to produce a binary image of the vasculature, b(i , j), for an image of size I x J. We want to achieve a robust segmentation of for a wide variety of images representing various states of the retinal disease.

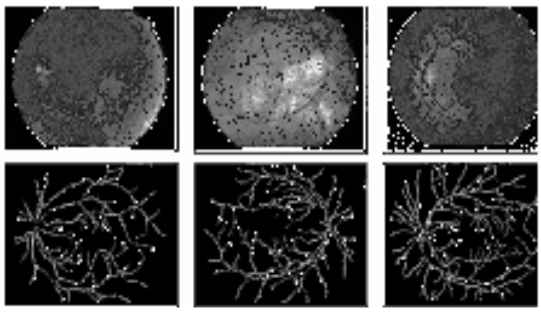


Fig.5 Vascular Segmentation with age-related macular degeneration (drusen)

Detection of Blood Vessels Boundaries
Algorithm 1 – Boundary detection using image statistics

Image statistics such as mean and standard deviation can be used in boundary detection. The algorithm called DBDED, which stands for decision-based directional edge detector uses image statistics. It is used to detect the boundaries of blood vessel tree in the retinal images. Each point that passes a local threshold is investigated as a an edge candidate. Then, the point (x , y) is to be a one-dimensional edge candidate from the east direction if it satisfies the following equation:

$$I(x,y) \geq Av[2I(x+1,y), I(x+2,y), I(x+3,y)] + Sd[2I(x+1,y), I(x+2,y), I(x+3,y)] + \eta$$

Where I(x, y) is the intensity of an image at coordinates x and y, η is a constant threshold and $Av(.)$ and $Sd(.)$ denote average and standard deviation operations respectively
The decision algorithm is applied and the points are considered as 2-D edges if they satisfy the following conditions
1.They are 1-D Edge candidates in at least two and at most seven directions.
2.If (x, y) is an edge candidate then, at least one of the immediate 8-neighboring points:
$\{(x+\beta, y+\gamma) \mid \beta \in (0,1,-1), \gamma \in (0,1,-1), \beta+\beta\gamma+\gamma \neq 0\}$
is also an edge candidate.

Algorithm 2 – Extraction of blood vessel boundaries using deformable models

The recent one of the methods of contour detection is deformable models are snake. A snake is an active contour model that is manually initiated near to the contour of interest. This contour model deforms according to some criteria and image features to finally stay to the actual contour(s) in the image. An energy function is formulated to obtain an estimate of the quality of the mode in terms of its internal shape, and external forces e.g. underlying image forces and user constraint forces. The energy function integrates the weighted linear combination of the internal and external forces of the contour.

Extraction of the core area of the blood vessel tree by tracing vessel centers

Algorithm : Extraction of blood vessel tree using the morphological reconstruction

Morphological reconstruction is to reconstruct an object in an image, called the marker image, containing at least one point belonging to that object from an image, called the mask image, containing that object and other objects and noise. An efficient implementation of morphological reconstruction can be described as follows:
1.Label the connected components of the mask image, (i.e.) each of these components is assigned a unique number.
2.Determine the labels of the connected components, which contain at least a pixel of the marker image.
3.Remove all the connected components that are not of the previous ones.

C.FEATURE EXTRACTION

The aim of texture analysis is texture recognition and texture based shape analysis. The texture can be studied in two levels namely statistical and structural. On the statistical level, the texture of an image is defined by a set of statistics extracted from the entire texture region. On the structural level, a texture is defined by sub patterns called primitives. The various statistical methods are based on capturing the variability in grey scale images. The textural character of an image depends on the spatial size of texture primitives. Large primitives give rise to coarse

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

texture (e.g. rock surface) and small primitives gives fine texture (e.g. silk surface).
Feature extraction can be done in two steps.

1. Features detecting optic nerve.
2. Features detecting diseases.

Features detecting optic nerve

Applying the pre-processing techniques like noise removal and histogram equalization, we obtain a better contrast image with the different objects having different textures. Each image is divided into large number of equal parts and local mean, standard deviation and variance are calculated. The characteristics of vessel structure are,

1.Retina luminancel(i,j) -This feature measures the brightness that can be helpful for locating the retinal lesions

$$\ell(i,j)=\frac{1}{M\cdot N}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} I(i-m,j-m)$$

which supports M X N region for every point (i , j) in the image.

2.Vessel density, ρ(i , j)-Vessel density is defined as the number of vessels existing in a unit area of the retina. Since the vasculature that feeds the retina enters the eye, the vessels tend to be most dense in this region.

$$\rho(i,j)=\frac{1}{M\cdot N}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} b_i(i-m,j-n),$$

which supports M X N region for every point (i , j) in the image.

3.Average vessel thickness, t(i , j)-Vessels are also observed to be thickest near the optic nerve since most branching of both the arterial and venous structures does not take place until the tree is more distal from the optic nerve.

$$t(i,j)=\frac{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} b(i-m,j-n)}{\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} b_i(i-m,j-n)},$$

which supports M X N region for every point (i , j) in the image.

4.Average Vessel Orientation, θ(i , j)-The vessels entering the eye are roughly perpendicular to the horizontal raphe of the retina. i.e. the demarcation line running through the optic nerve and fovea. The result is an observation of vascular orientation being ±90 ° relative to the horizontal raphe when entering the eye and becoming more parallel (i.e., 0°) as the distance from the optic nerve increases.

$$\theta(i,j)=\frac{1}{M\cdot N}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} b_i(i-m,j-n)\cdot\cos\theta(i-m,j-n),$$

which supports M X N region for every point (i , j) in the image.

Features detecting retinal diseases

Statistical texture features describe texture in a form, which is suitable for pattern recognition. As a result each texture is described by a feature vector of properties, which represents a point in a multi dimensional feature space.

a) Mean-The nth moment of about the mean is

$$\mu_n(z) = \sum_{i=0}^{L-1} (z_i - m)^n \, p(z_i) \qquad (1)$$

where m is the **mean value** of z (the average gray level) :

$$m = \sum_{i=0}^{L-1} z_i \, p(z_i)0$$

Note from Eq. (1) that $\mu_0 = 1$ and $\mu_1 = 0$.

b) Variance-The second moment [the **variance** $\sigma^2(z) = \mu_2(z)$] is of particular importance in texture description. It is a measure of gray-level contrast that can be used to establish descriptors of relative smoothness.

c) Skewness- The third moment,

$$\mu_3(z) = \sum_{i=0}^{L-1} (zi - m)^3 \, p(z_i)$$

is a measure of the **skewness** of the histogram.

d) Entropy

$H(\mathbf{z}) = -\int p(\mathbf{z}) \ln p(\mathbf{z}) \, d\mathbf{z}$ is Shannon's entropy of the image window **z**, and *p* is the distribution of the grey levels in the considered window. We approximate the entropy as:

$$H(\mathbf{z}) \approx -\frac{1}{N_z}\sum_{z_i\in\mathbf{z}} \ln\frac{1}{N_p}\sum_{z_j\in W_p} g_\psi(z_i - z_j)$$

e) Correlation distance-The distances among all the images can be computed using correlation distance.

$$d_{rs} = 1 - \frac{(x_r - \bar{x}_r)(x_s - \bar{x}_s)^T}{\left[(x_r - \bar{x}_r)(x_r - \bar{x}_r)^T\right]^{\frac{1}{2}}\left[(x_s - \bar{x}_s)(x_s - \bar{x}_s)^T\right]^{\frac{1}{2}}},$$

where $\bar{x}_r = \frac{1}{n}\sum_j x_{rj}$ and $\bar{x}_s = \frac{1}{n}\sum_j x_{sj}$.

g) Zernike moments (ZM)- The complex zernike moments of order n with repetition *l* are defined as

$$A_{nl} = \frac{n+1}{\pi}\int_0^{2\pi}\int_0^{\infty} [V_{nl}(r,\theta)]^* \cdot f(r\cos\theta, r\sin\theta) r \, dr \, d\theta$$

Where n = 0,1,2,…., and l takes on positive and negative integer values subject to the conditions $n - |l| = \text{even}, \quad |l| \le n.$ The symbol * denotes complex conjugate. The zernike polynomials

$$v_{nl}(x,y) = V_{nl}(r\cos\theta, r\sin\theta) = R_{nl}(r)e^{il\theta}$$

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

are a complete set of complex-valued functions orthogonal to the unit disk.

The function f(x, y) can be expanded in terms of zernike polynomials over the unit disk as

$$f(x, y) = \sum_{n=0}^{\infty} \sum_{\substack{l=-\infty \\ n-|l|=\text{even} \\ |l| \le n}}^{\infty} A_{nl} V_{nl}(x, y)$$

D.Classification of retinal disease using Artificial Neural Networks

Artificial Neural Networks (ANN) has been used in a number of different ways in medicine and medically related fields. The principal advantage of ANN is to generalize, adapting to signal distortion and noise without the loss of robustness. Auto Associative Neural Network (AANN) is a network having the same number of neurons in input and output layers, and the less in the hidden layers. The network is trained using the input vector itself as the desired output. This training leads to organize a compression/encoding network between the input layer and the hidden layer, and a decoding layer between the hidden layer and output layer. Each of the autoassociative networks is trained independently for each class using the feature vector of the class. The squared error between an input and the output is generally minimized by the network of the class to which the input pattern belongs. This property enables us to classify an unknown input pattern. The unknown pattern is fed to all the networks, and is classified to the class with minimum squared error.

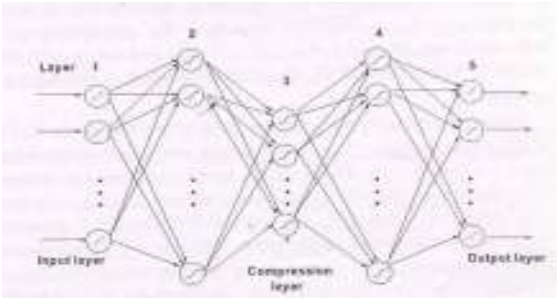


Fig 5 Auto Associative Neural Network

IV.CONCLUSION

Computer aided diagnosis of retinal diseases is one of the most important tasks when dealing with a huge population. This paper gives a survey of the classical and up-to-date methods for classifying and diagnosing the type of retinal disease and detecting its features after diagnosis at an earlier stage of the disease. Although a lot of work has been done, automatic diagnosis of retinal diseases at an earlier stage still remains an open problem.

The paper gives only the frame work for diagnosing human retinal diseases. This can be implemented using matlab. Each module can be tested individually with a test data of size 100. The results can be classified into four phases: true positive, true negative, false positive, false negative. The major goal of the paper is to provide a comprehensive reference source for the researchers involved in automatic diagnosis of retinal images. This framework can be extended to any number of retinal diseases in future.

V.REFERENCES

[1] S. Swarna Parvathi, N. Devi, (2007), "Automatic Drusen Detection from Colour Retinal Images," IEEE Proceedings of the International Conference on “Computational Intelligence and Multimedia Applications”, Mepco, vol 2, pp: 377-381

[2] Nguyen.H., Roychoudhry.A., Shannon.A, (1997), “Classification of diabetic retinopathy lesions from stereoscopicfundus images,” IEEE Proceedings of the 19th Annual International Conference on Engineering in Medicine and Biology Society, Vol 1, Issue 30 Oct-2 1997, Page(s):426 – 428.

[3] Usman Akram. M. , Sarwat Nasir, Almas Anjum.M., Younus Javed.M., (2009) “Background and noise extraction from colored retinal images,” IEEE Proceedings of the WRI World Congress on Computer Science and Information Engineering, vol.6, March 31- April 2 2009, pp.573-577.

[4] Yitao Liang, Lianlian He, Chao Fan, Feng Wang, Wei Li, (2008), “PreProcessing Study of Retinal Image Based On Component Extraction,” IEEE Proceedings of International Symposium on IT in Medicine and Education, 12-14 Dec. 2008, pages: 670-672.

[5] Mai S. Mabrouk, Nahed H. Solouma, Yassad M.kadah, (2006), “Survey of retinal image registration and segmentation,” icgst, GVIP Journal, vol 6, Issue 2, September 2006.

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

Changing Neighbors k-Secure Sum Protocol for Secure Multi-Party Computation

Rashid Sheikh, Beerendra Kumar
SSSIST, Sehore, INDIA
,
.

Durgesh Kumar Mishra
Acropolis Institute of Technology and Research
Indore, INDIA .

***Abstract*- Secure sum computation of private data inputs is an important component of Secure Multi-party Computation (SMC).In this paper we provide a protocol to compute the sum of individual data inputs with zero probability of data leakage. In our proposed protocol we break input of each party into number of segments and change the arrangement of the parties such that in each round of the computation the neighbors are changed. In this protocol it becomes impossible for semi honest parties to know the private data of some other party.**

***Keywords*- Secure Multi-party Computation (SMC), Privacy, Computation Complexity, Semi honest Parties, k-Secure Sum Protocol, Information Security, Trusted Third Party (TTP).**

I. INTRODUCTION

In today's world of information technology opportunities exist for joint computation requiring privacy of the inputs. These computations occur between parties which may not have trust in one another. In literature this subject is called Secure Multi-party Computation (SMC). It is aimed at privacy of individual inputs and the correctness of the result. Formally in SMC the parties P_1, P_2 ,..., P_n want to compute some common function $f(x_1, x_2, \ldots, x_n)$ of inputs $x_1, x_2, \ldots, x_n$ such a party can know only its own input x_i and the value of the function f. The SMC problems use two computation models; ideal model and real model. In ideal model there exists a Trusted Third Party (TTP) which accepts inputs from all the parties, evaluates the common function. In real model the parties agree on some protocol which allows all the parties to evaluate the function. For example if two banks cooperatively want to know details about some customer but no bank is willing to disclose the details of the customer to other bank due to privacy of customer or policy of the bank. In such situations the SMC solutions are important. The best example of SMC is the secure sum computation where all the cooperating parties want to compute the sum of their individual data inputs while preserving confidentiality of inputs [10]. The secure sum protocol proposed by Clifton *et al.* in [10] uses randomization method for computing the sum. In this protocol two adjacent parties to a middle party can collude maliciously to know the data of a middle party. We proposed new protocols in [11] where the probability of data leakage has been reduced by segmenting the data block into a fixed number of segments.

In this paper we propose a novel secure sum computation protocol with zero probability of data leakage. In this protocol we change the position of the parties so that the neighbors are changed in each round of the computation. This protocol which is an extension of our previous protocol we call as *ck-Secure Sum Protocol.*

2. BACKGROUND

The subject of SMC began in 1982 when Yao proposed his millionaire's problem in which two millionaires wanted to know who was richer without revealing individual wealth to each other [1]. The solution provided was for semi honest. The semi honest parties follow the protocol but also try to know some other information. The concept was extended to multi-party computation [2]. Goldreich *et al.* also used circuit evaluation protocols for secure computation. Some real life applications of SMC emerged like Private Information Retrieval (PIR) [3, 4], Privacy-preserving data mining [5, 6], Privacy-preserving geometric computation [7], Privacy-preserving scientific computation [8], Privacy-preserving statistical analysis [9] etc. An excellent review of SMC is provided by Du *et al.* in [12] where they developed a framework for SMC problem discovery and transformation of normal problem to SMC problem. A review of SMC problems with a focus on telecommunication systems is provided by Oleshchuk *et al.* in [13]. Anonymity

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

enabled SMC was proposed by Mishra *et al.* in [14] where the identities of the parties are hidden for achieving privacy.

In this paper the protocol is motivated by the work of Clifton *et al.* [10] where they proposed a toolkit of components for solution to SMC problems. They proposed that one of components of the toolkit for SMC is the secure sum computation. Secure sum computation is used in many distributed data mining applications where many geographically distributed sites compute sum of values from individual sites. The secure sum protocol proposed in [10] used random numbers for privacy of individual data inputs. In this scheme any two parties P_{i-1} and P_{i+1} can collude to know the secret data of party P_i by performing only one computation. We proposed *k-Secure Sum Protocol* and *Extended k-Secure Sum Protocol* in [11] where the probability of data leakage is significantly reduced by breaking the data block of individual party in number of segments. The probability of data leakage decreases as the number of segments in a data block is increased.

As per our survey no secure sum protocol is available in the literature with zero probability of data leakage when two neighbors collude. In this paper we proposed zero probability protocol for secure sum computation namely *ck-Secure Sum Protocol* in which neighbors are changed in each round of computation.

3. Proposed Architecture and the Protocol Description

The initial architecture of the protocol is shown in fig 1 where parties are arranged in a ring. Each party breaks the data block into *k* segments which is equal to *n*-1. For example in fig 1 four parties break their data block into three segments. Initially the parties are arranged sequentially as $P_1, P_2, ..., P_n$. In the next round of the computation P_2 exchanges its position with P_3 and in subsequent rounds P_2 exchanges its position with P_4 and so on until P_n is reached.

3.1 Informal Description of ck-Secure Sum Protocol

We observed in secure sum protocol [10] and k-secure sum protocol [11] that a middle party can be hacked by two neighbor parties with some probability. The motivation for *ck-Secure Sum Protocol* is that we change the neighbors in each round of segment computation. Thus it is guaranteed that no two semi honest parties can learn all the data segments of a victim party. In this protocol also each party breaks the data

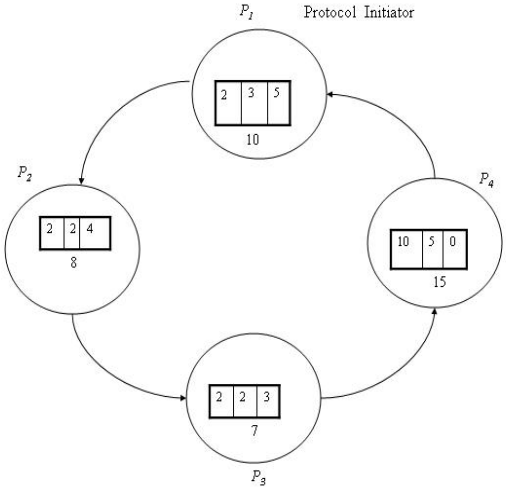


Figure 1: Initial architecture of *ck-Secure Sum Protocol*

block into $k = n-1$ segments where n is the number of parties involved in the cooperative sum computation. We propose P_1 to be the protocol initiator. The position of the protocol initiator is kept fixed in each round of computation. For the first round of the computation parties are arranged in a serial fashion as $P_1, P_2, ..., P_n$. The protocol initiator starts computation using k-secure sum protocol to get the sum of first segment of each party. Before second round of computation starts P_2 exchanges its position with P_3. In next round of the computation P_2 exchanges its position with P_4 and so on until P_2 exchanges its position with P_n. Generalizing the method we can say that in i^{th} round of the computation P_2 exchanges its position with P_{i+1} until P_n is reached. In each round of computation segments are added using k-secure sum protocol [11] and the partial sum is passed to the next party until all the segments are added and the sum is announced by the protocol initiator party. Snapshots for a four-party case are shown in fig 2

3.2 Formal Description of ck-Secure Sum Protocol

The *ck-Secure Sum Protocol* is an extension of *k-Secure Sum Protocol* [11] and is based on changing neighbors in each round of segment computation. The party P_1 is selected as the protocol initiator party which starts the computation by sending the first data segment. The party traverses towards P_n in each round of the computation. The number of parties for this

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

protocol must be four or more. When all the rounds of segment summation is completed the sum is announced by the protocol initiator party

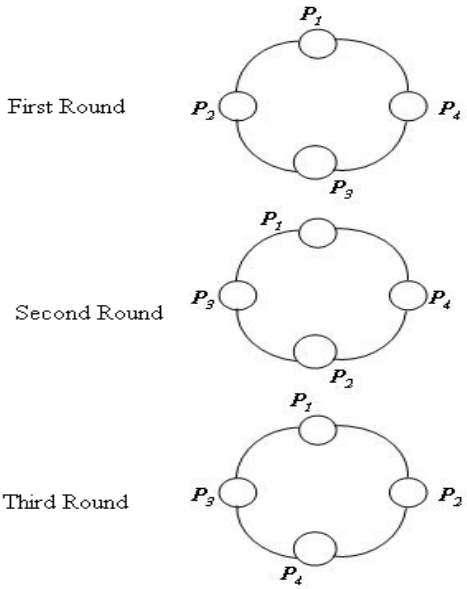


Figure 2: Snapshots of *ck-Secure Sum Protocol* for four-party case.

The algorithm: *ck-Secure Sum*

1. Define P_1, P_2, ..., P_n as n parties where $n >= 4$.
2. Assume these parties have secret inputs x_1, x_2, ..., x_n .
3. Each party P_i breaks its data x_i into $k= n-1$ segments d_{i1} , d_{i2} , ..., d_{ik} such that $\sum d_{ij} = x_i$ for j =1 to k.
4. Arrange parties in a ring as P_1, P_2 , ..., P_n and select P_1 as the protocol initiator.
5. Assume $rc = k$ and $S_{ij} = 0$. /* S_{ij} *is partial sum and rc is round counter*/*
6. While rc!=0
   begin
   for j = 1 to k
   begin
   for i =1 to n
   begin
   starting from P_1 each party computes cumulative sum S_{ij} of its next segment and thereceived sum from its neighbor and sends to the next party in the ring
   end
   P_2 exchanges its position with $P_{(j+2) \bmod n}$
   end
   $rc = rc - 1$
   end
7. Party P_1 announces the result as S_{ij} .
8. End of algorithm.

3.3 Performance analysis of ck-Secure Sum Protocol

In this protocol each data segment is secret of the party and chosen with its own way. If two neighbor parties collude they can know only one segment in one round of the computation. The protocol guarantees that a party will not have same two neighbors in all the rounds of the computations. The neighbors are changed at least once during secure sum computation. Thus any two neighbors to a middle party cannot know all the segments of a party. The semi honest parties cannot learn more information than the result. Thus the probability of data leakage by two colluder parties to a middle party is zero. Number of rounds of computation is n-1 and the number of exchanges between parties is n-2. The only drawback of this scheme is that the topology of the computational network changes in each round of the computation. The communication and computation complexity both are $O(n^2)$.

3.4 Conclusion and future scope

Secure sum computation is an important element of toolkit for SMC solution. Protocols are needed for secure sum computation with greater security to individual data. The protocol *ck-Secure Sum Protocol* changes neighbors in each round of computation. Our proposed protocol provides zero probability of data leakage by two colluding parties when they want to attack data of a middle party. This is an appreciable improvement over previous protocols available in the literature. Efforts can be made to reduce the computation and the communication complexity preserving the property of zero hacking.

References

[1] A.C.Yao, "protocol for secure computations," in *proceedings of the 23rd annual IEEE symposium on foundation of computer science*, pages 160-164, Nov.1982.

[2] O. Goldreich, S. Micali, and A. Wigderson, "How to play any mental game," in *STOC '87: Proceedings of the nineteenth annual ACM conference on Theory of computing*. New York, NY, USA: ACM, pages 218-229 1987.

[3] B.Chor and N.Gilbao. "Computationally Private Information Retrieval (Extended Abstract)," In *proceedings of 29th annual ACM Symposium on Theory of Computing*, El Paso, TX USA, May 1997.

[4] B. Chor, E. Kushilevitz, O. Goldreich,and M. Sudan, "Private Information Retrieval ,"

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

In *proceedings of the 36th Annual IEEE Symposium on Foundations of Computer Science*, Milwaukee WI, pages 41-50, Oct. 1995.

[5] Y. Lindell and b. Pinkas, "Privacy preserving data mining," in *advances in cryptography-*Crypto2000, *lecture notes in computer science, vol. 1880*, 2000.

[6] R. Agrawal and R. Srikant. "Privacy-Preserving Data Mining," In *proceedings of the 2000 ACM SIGMOD on management of data*, Dallas, TX USA, pages 439-450, May 15-18 2000.

[7] M. J. Atallah and W. Du. "Secure Multiparty Computational Geometry," *In proceedings of Seventh International Workshop on Algorithms and Data Structures(WADS2001)*. Providence, Rhode Island, USA, pages 165-179, Aug. 8-10 2001.

[8] W. Du and M.J. Atallah. "Privacy-Preserving Cooperative Scientific Computations," *In 14th IEEE Computer Security Foundations Workshop*, Nova Scotia, Canada, pages 273-282, Jun. 11-13 2001.

[9] W. Du and M.J.Atallah, "Privacy-Preserving Statistical Analysis," *In proceedings of the 17th Annual Computer Security Applications Conference*, New Orleans, Louisiana, USA, pages 102-110, Dec. 10-14 2001.

[10] C. Clifton, M. Kantarcioglu, J.Vaidya, X. Lin, and M. Y. Zhu, "Tools for Privacy-Preserving Distributed Data Mining,"*J. SIGKDD Explorations, Newsletter, vol.4, no.2, ACM Press*, pages 28-34, Dec. 2002.

[11] R. Sheikh, B. Kumar and D. K. Mishra, "Privacy-Preserving k-Secure Sum Protocol," in *International Journal of Computer Science and Information Security*, *vol. 6 no.2*, pages 184-188, Nov. 2009.

[12] W. Du and M.J. Atallah, "Secure Multiparty Computation Problems and Their Applications: A Review and Open Problems," In *proceedings of new security paradigm workshop*, Cloudcroft, New Maxico, USA, pages 11-20, Sep. 11-13 2001.

[13] V. Oleshchuk, and V. Zadorozhny, "Secure Multi-Party Computations and Privacy Preservation:Results and Open Problems," *Telektronikk: Telenor's Journal of Technology, vol. 103, no.2*, 2007.

[14] D. K. Mishra, M. Chandwani, "Extended Protocol for Secure Multiparty Computation using Ambiguous Identity," *WSEAS Transaction on Computer Research, vol. 2, issue 2*, Feb. 2007.

Authors Profile

Dr. Durgesh Kumar Mishra
Ph - +91 9826047547, +91-731-4730038
Email: durgeshmishra@ieee.org

Dr. Durgesh Kumar Mishra has received M.Tech. in Computer Science from DAVV, Indore in 1994 and PhD in Computer Engineering in 2008. Presently he is working as Professor (CSE) and Dean (R&D) in Acropolis Institute of Technology & Research, Indore, MP, India. He is having around 20 Yrs of teaching experience and more than 5 Yrs of research experience. He has completed his research work with Dr. M. Chandwani, Director, IET-DAVV Indore, MP, India on Secure Multi-Party Computation. He has published more than 65 papers in refereed International/National Journals and Conferences including IEEE and ACM He is a senior member of IEEE and Secretary of IEEE MP-Subsection under the Bombay Section, India. Dr. Mishra has delivered tutorials in IEEE International conferences in India as well as other countries. He is the programme committee member of several International conferences. He visited and delivered invited talks in Taiwan, Bangladesh, USA, UK etc. on Secure Multi-Party Computation of Information Security. He is an author of one book. He is reviewer of three international journals of information security. He is Chief Editor of *Journal of Technology and Engineering Sciences*. He has been a consultant to industries and Government organization like Sales tax and Labor Department of Government of Madhya Pradesh, India.

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

Rashid Sheikh
Ph. +91 9826024087
Email: rashidsheikhmrsc@yahoo.com

Rashid sheikh has received his Bachelor of Engineering degree in Electronics and TelecommunicationEngineering from Shri Govindram Seksaria Institute of Technology and Science, Indore, M.P., India in 1994. He has 15 years of teaching experience. His subjects of interest include Computer Architecture, Computer Networking, Electrical Circuit analysis, Digital Computer Electronics, Operating Systems and Assembly Language Programming. Presently he is pursuing M. Tech. (Computer Science and Engineering) at SSSIST, Sehore, M.P., India. He has published four research papers in National Conferences and one research paper in international journal. His research areas are Secure Multiparty Computation and MANET. He is the author of ten books on Computer Organization and Architecture.

Beerendra Kumar
Ph. +91 9770435336
Email: beerucsit@gmail.com

Beerendra Kumar has received B.Tech. (Bachelor of Technology) degree in Computer Science and Information Technology from Institute of Engineering and Technology, Rohilkhand University, Bareilly (U.P), India in 2006. He has completed his M.Tech. (Master of Technology) in Computer Science from SCS, Devi Ahilya University, Indore, India in 2008.He has two years of teaching experience. His subjects of interest include Computer Networking, Theory of Computer Science, Data Mining, Operating Systems and Analysis & Design of Algorithms. He has published three research papers in national conferences and one research paper in international journal. His research areas are Computer Networks, Data Mining, Secure Multiparty Computations and Neural Networks.

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

A Probabilistic Model For Sequence Analysis

Amrita Priyam
Dept. of Computer Science and Engineering
Birla Institute of Technology
Ranchi, India.

B. M. Karan[+], G. Sahoo[++]
[+]Dept. of Electronics and Electrical Engineering
[++]Dept. of Information Technology
Birla Institute of Technology
Ranchi, India

Abstract— **This paper presents a probabilistic approach for DNA sequence analysis. A DNA sequence consists of an arrangement of the four nucleotides A, C, T and G and different representation schemes are presented according to a probability measure associated with them. There are different ways that probability can be associated with the DNA sequence: one way is when the probability of an occurrence of a letter does not depend on the previous one (termed as unsuccessive probability) and in another scheme the probability of occurrence of a letter depends on its previous letter (termed as successive probability). Further, based on these probability measures graphical representations of the schemes are also presented. Using the digram probability measure one can easily calculate an associated probability measure which can serve as a parameter to check how close is a new sequence to already existing ones.**

Keywords-Successive Probability; Unsuccessive Probability; Transition Probability; Digram Probability;

I. INTRODUCTION

A DNA sequence is a succession of the letters A, C, T and G. The sequences are any combination of these letters. A physical or mathematical model of a system produces a sequence of symbols according to a certain probability associated with them. This is known as a stochastic process, that is, it is a mathematical model for a biological system which is governed by a set of probability measure. The occurrence of the letters can lead us to the further study of genetic disorder. The stochastic process is also known mathematically as Discrete Markov Process. There are different ways to use probabilities for depicting the DNA sequences. One scheme is where each occurrence of a letter is independent of the occurrence of any other letter. That is, the probability of occurrence of a letter does not depend on the occurrence of the previous one. In yet another scheme, the occurrence of a letter depends on the occurrence of the previous one. The earlier one is termed in as unsuccessive probability and the later one is termed as successive probability. From this study we can further show the relationship between sequence and genetic variations. It can also lead to a more powerful test for identifying particular classes of genes or proteins which has been illustrated by an example.

II. DNA SEQUENCE

DNA sequences is a succession of letters representing the primary structure of a real or hypothetical DNA molecule or strand, with the capacity to carry information as described by the central dogma of molecular biology. There are 4 nucleotide bases (A – Adenine, C – Cytosine, G – Guanine, T – Thymine). DNA sequencing is the process of determining the exact order of the bases A, T, C and G in a piece of DNA. In essence, the DNA is used as a template to generate a set of fragments that differ in length from each other by a single base. The fragments are then separated by size, and the bases at the end are identified, recreating the original sequence of the DNA. The most commonly used method of sequencing DNA the dideoxy or chain termination method was developed by Fred Sanger in 1977 (for which he won his second Nobel Prize). The key to the method is the use of modified bases called dideoxy bases; when a piece of DNA is being replicated and a dideoxy base is incorporated into the new chain, it stops the replication reaction.

Most DNA sequencing is carried out using the chain termination method. This involves the synthesis of new DNA strands on a single standard template and the random incorporation of chain-terminating nucleotide analogues. The chain termination method produces a set of DNA molecules differing in length by one nucleotide. The last base in each molecule can be identified by way of a unique label. Separation of these DNA molecules according to size places them in correct order to read off the sequence.

III. DIFFERENT PROBABILISTIC APPROACHES FOR SEQUENCE REPRESENTATION

A DNA sequence is essentially represented as a string of four characters A, C, T, G and looks something like ACCTGACCTTACG. These strings can also be represented in terms of some probability measures and using these measures it can depicted graphically as well. This graphical representation matches the Markov Hidden Model. Some of these schemes are presented in this paper. A physical or mathematical model of a system produces a sequence of symbols according to a certain probability associated with them. This is known as a stochastic process. There are

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

different ways to use probabilities for depicting the DNA sequences

A. *Unsuccessive Probability*

In this representation scheme the probability of the next occurrence of a letter does not depend on the previous letter. There can be two different representation schemes in this: one which uses equal probability for each letter and another which assumes a fixed probability for each letter.

- *Equal Probability* - Suppose we have a 4 letter code consisting of the 4 letters A, C, T, G which can be chosen each with equal probability 0.25, successive choices being independent. This would lead to a sequence of which the following is a typical example

  ACCATGGACTTAGCTACTGG

- *Unequal Probability* - For a DNA sequence series of length 20, where each letter has a probability of 0.3, 0.2, 0.3, 0.2 respectively, with successive choices are independent. A typical message from this source is:

  ACTTGAAATTCGGACCTGAT

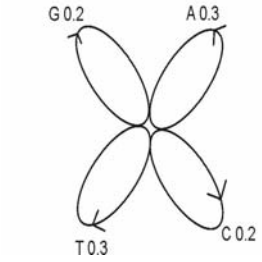


Fig 1: Graphical Representation of Unequal Probability Sequence

B. *Successive Probability*

This section presents two schemes for representing a DNA sequence: one where successive letter depends on the preceding one and in another scheme the letters are used to form words and a probability is associated with each word.

- *Letter Probability* - A more complicated structure is obtained if successive symbols are not chosen independently but their probabilities depend on preceding letters. In the simplest case of this type a choice depends only on the preceding letter and not on ones before that. The statistical structure can then be described by a set of transition probabilities $p_i(j)$, the probability that a letter i is followed by letter j. The indices i and j range over all the possible symbols. A second equivalent way of specifying the structure is to give the "digram" probabilities $p(i,j)$, i.e., the relative frequency of the digram $i\ j$. The letter frequencies $p(i)$, (the probability of letter i), the transition probabilities $p_i(j)$ and the digram probabilities $p(i,j)$ are related by the following formulas:

$$p(i) = \sum_j p(i,j) = \sum_j p(j,i) = \sum_j p(j)p_j$$

$$p(i,j) = p(i)p_i(j)$$

$$\sum_j p_i(j) = \sum_i p(i) = \sum_{i,j} p(i,j) = 1$$

  In the example presented in this paper we have used the letter probability for representing the sequence.

- *Word Probability* - A process can also be defined which produces a text consisting of a sequence of words. Since a DNA sequence consists of typically 4 letters A, C, T, G any word could be made up of these four letters. For example the typical DNA sequence may consist of any combination of the following words:
  ACTG TACG ACGT AATC AGTG TCCA CAAG CCTG

  This can be depicted graphically as in figure 2.

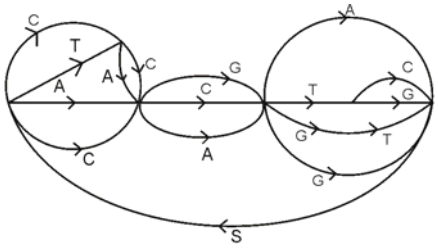


Fig 2: Graphical Representation of Word Probability Sequence

The various probability (transition and digram) are calculated according to the algorithm given below:

ALGORITHM

SERIES (DATA, N, LOC): Here DATA is a linear array with N elements and LOC acts as a pointer that keeps the record of occurrence of the nucleotide.

1. Initialize LOC = 0 and initialize all variables to zero.

2. Repeat while LOC ≠ N,
   IF DATA [LOC] = 'A'
   Set A = A+ 1
   IF DATA [LOC +1] = 'A' SET AA = AA+1

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

IF DATA [LOC +1] = 'C' SET AC = AC+1
IF DATA [LOC +1] = 'T' SET AT = AT+1
IF DATA [LOC +1] = ' G' SET AG = AG +1

Else IF DATA [LOC] = ' C'
Set C = C+1
IF DATA [LOC + 1] = 'A' SET CA = CA+1
IF DATA [LOC +1] = ' C' SET CC = CC+1
IF DATA [LOC +1] = ' T' SET CT = CT+1
IF DATA [LOC +1] = ' G' SET CG = CG +1

Else IF DATA [LOC] = 'T'
Set T = T+1
IF [LOC +1] = 'A' SET TA = TA+ 1
IF [LOC +1] = 'C' SET TC = TC+1
IF [LOC + 1] = ' T' SET TT = TT + 1
IF [LOC + 1] = ' G' SET TG = TG +1

Else IF DATA [LOC] = ' G'
Set G = G+1
IF DATA [LOC +1] = 'A' SET GA = GA +1
IF DATA [LOC +1] = 'C' SET GC = GC +1
IF DATA [LOC +1] = 'T' SET GT = GT +1
IF DATA [LOC+1] = 'G' SET GG = GG+1

3. SET A' = A/N , C' = C/N , G' = G/N, T' = T / N
AA' = AA* A'; AC' = AC*A'; AT ' = AT* A';
AG ' =AG* A'; CA' = CC* C' ; CC' = CC*C';
CT' = CT * C'; CG' = CG * C'; TA' = TA * T';
TC' = TC * T'; TT' =TT * T'; TG' = TG* T';
GA' = GA *G'; GC' = GC *G'; GT' = GT * G';
GG' = GG * G';

4. End

LETTER FREQUENCY

(i)	*P(i)*
A	No of A'S / Size of string = (A' Say)
C	No of C'S / Size of string = (C' Say)
T	No of T'S / Size of string = (T' Say)
G	No of G'S / Size of string = (G' say)

DIGRAM PROBABILITY

P(i, j)	A	C	T	G
A	No. of AA * A' = AA'	No. of AC * A' = AC'	No. of AT * A' = AT'	No. of AG * A' = AG'
C	No. of CA * C' = CA'	No. of CC * C' = CC'	No. of CT * C' = CT'	No. of CG * C' = CG'
T	No. of TA * T' = TA'	No. of TC * C' = TC'	No. of TT * T' = TT'	No. of TG * T' = TG'
G	No. of GA * G' = GA'	No. of GC * C' = GC'	No. of GT * T' = GT'	No. of GG * G' = GG'

TRANSITION PROBABILITY

$P_i(j)$	A	C	T	G
A	A' * AA'	A' * AC'	A' * AT'	A' * AG'
C	C' * CA'	C' * CC'	C' * CT'	C' * CG'
T	T' * TA'	T' * TC'	T' * TT'	T' * TG'
G	G' * GA'	G' * GC'	G' * GT'	G' * GG'

On using the algorithm on the different H1N1 viruses we get the following transition probability tables:

Type 1:

TABLE I

	A	C	T	G
A	0.13	0.06	0.09	0.08
C	0.09	0.04	0.05	0.02
T	0.06	0.05	0.05	0.07
G	0.08	0.04	0.05	0.06

Type 2:

TABLE II

	A	C	T	G
A	0.13	0.06	0.09	0.08
C	0.08	0.04	0.04	0.02
T	0.06	0.05	0.06	0.07
G	0.08	0.04	0.04	0.06

Type 3:

TABLE III

	A	C	T	G
A	0.08	0.05	0.08	0.08
C	0.08	0.03	0.06	0.02
T	0.05	0.06	0.06	0.09
G	0.08	0.06	0.04	0.07

A graphical representation for the type 1 H1N1 virus can be constructed based on the digram probability as follows:

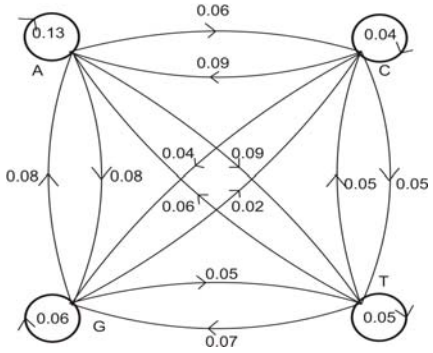


Fig 3: Graphical Representation of Type 1 H1N1 Sequence

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

Now if any new sequence comes which has to be categorized as any one of the three types we can use the transition probability table of the existing types and calculate the probability measure. For example consider a new sequence ***tggtgctggc***

Probability for Type 1: 0.07 * 0.06 * 0.05 * 0.07 * 0.04 * 0.05 * 0.07 * 0.06 * 0.04 = 4.9 X 10^{-12}

Probability for Type 2: 0.06 * 0.06 * 0.04 * 0.07 * 0.04 * 0.04 * 0.07 * 0.06 * 0.04 = 2.7 X 10^{-11}

Probability for Type 3: 0.09 * 0.07 * 0.04 * 0.09 * 0.06 * 0.06 * 0.09 * 0.07 * 0.06 = 3 X 10^{-11}

Since the probability measure for type 3 comes out to be the greatest we can conclude that the chance for the new DNA sample to be the type 3 virus is more.

REFERENCES

[1] C. E. Shanon, "A Mathematical Theory of Communciation".

[2] T. Dewey and M. Herzel, "Application of Information Theory to Biology".

[3] D. W. Mount, " Bioinformatics, Sequence and Genome Analysis", 2nd edition, CBS publishers and Distributors.

[4] W. J. Ewens, G. R. Grant, "Statistical Methods in Bioinformatics: AN Introduction", Vol. 13, 2nd edition, Springer.

[5] L. R. Rabenir, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of the IEEE, Vol. 77, No. 2, Feb 1989.

[6] J. L. Massey, "Applied Digital Information Theory", Lecture Notes.

[7] K. Senthamaraikannan, D. Nagarajan, V. Nagaranjan, "Transition Probability Matrix of Mothers and Newborn Heamoglobin Count in Kanyakumaro District", Ethno-Med, 57 – 59, 2008.

[8] I. Csiszar, P. C. Shields, "Information Theory Statistics: A Tutorial", Foundations and Trends in Communication and Information Theory,vol 1, No. 4, 2004.

[9] Po-Ning Chen, Fady Alajaji, " Lecture Notes in Information Theory", Vol 2.

[10] Andras Telcs, "Transition Probability Estimates for Reversible Markov Chains", Electronics Communication In Probability, 5, 29 – 37, 2000.

[11] Y. Ephraim, L. R. Rabiner, "On the Relations Between Modeling Approaches for Speech Recognition", IEEE transactions on Information Theory, vol. 36, no. 2, March 1990.

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

Dual Watermarking Scheme with Encryption

R.Dhanalakshmi
PG Scholar
Dept of CSE
Sri Venkateswara college
Of Engineering
Post Bag No. 3, Pennalur
Sriperumbudur
602 105
India

K.Thaiyalnayaki
Assistant Professor
Dept of IT
Sri Venkateswara college
of Engineering
Post Bag No. 3, Pennalur
Sriperumbudur
602 105
India

Abstract- *Digital Watermarking is used for copyright protection and authentication. In the proposed system, a Dual Watermarking Scheme based on DWT-SVD with chaos encryption algorithm, will be developed to improve the robustness and protection along with security. DWT and SVD have been used as a mathematical tool to embed watermark in the image. Two watermarks are embedded in the host image. The secondary is embedded into primary watermark and the resultant watermarked image is encrypted using chaos based logistic map. This provides an efficient and secure way for image encryption and transmission. The watermarked image is decrypted and a reliable watermark extraction scheme is developed for the extraction of the primary as well as secondary watermark from the distorted image.*

I. INTRODUCTION

The process of embedding information into another object/signal can be termed as watermarking. Watermarking is mainly used for copy protection and copyright-protection. Historically, watermarking has been used to send ``sensitive'' information hidden in another signal. Watermarking has its applications in image/video copyright protection. The characteristics of a watermarking algorithm are normally tied to the application it was designed for. The following explain the requirements of watermarking: i) Imperceptibility - A watermark is called perceptible if its presence in the marked signal is noticeable, but non-intrusive. A watermark is called imperceptible if the cover signal and marked signal are indistinguishable with respect to an appropriate perceptual metric.

ii) Robustness - The watermark should be able to survive any reasonable processing inflicted on the carrier. A watermark is called fragile if it fails to be detected after the slightest modification. Fragile watermarks are commonly used for tamper detection (integrity proof).

iii) Security - The watermarked image should not reveal any clues of the presence of the watermark, with respect to un-authorized detection, or undetectability or unsuspicious.

A visible watermark howsoever robust it may be can always be tampered using various software. To detect such kind of tampering (in worst case to protect the image when the visible watermark is fully removed) an invisible watermark can be used as a back up. The dual watermark is a combination of a visible watermark and an invisible watermark. The visible watermark is first inserted in the original image and then an invisible watermark is added to the already visible-watermarked image. The final watermarked image is the dual watermarked image.

The first applications were related to copyright protection of digital media. In the past duplicating artwork was quite complicated and required a high level of expertise for the counterfeit to look like the original. However, in the digital world this is not true. Now it is possible for almost anyone to duplicate or manipulate digital data and not lose data quality. Similar to the process when artists creatively signed their paintings with a brush to claim copyrights, artists of today can watermark their work by hiding their name within the image. Hence, the embedded watermark permits identification of the owner of the work.

With the growing threat of piracy in the Internet and copyright infringement cases, watermarks are sure to serve an important role in the future of intellectual property rights.

II. PROPOSED SYSTEM

Dual watermarking scheme based on DWT and Singular Value Decomposition (SVD) along with the chaos based encryption technique is proposed. After decomposing the cover image into four bands (LL, HL, LH, and HH), we apply the SVD to each band, and modify the singular values of the cover image with the singular values of the watermarked primary watermark. When the primary watermark image is in question, the invisible secondary watermark can provide rightful ownership. Modification in all frequencies allows the development of a watermarking scheme that is

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

robust to a wide range of attacks. SVD transform is performed on all the images and sum up the singular values to find the new singular values. Both the watermarks are embedded in the same manner and the watermarked primary watermark is encrypted using chaos encryption.

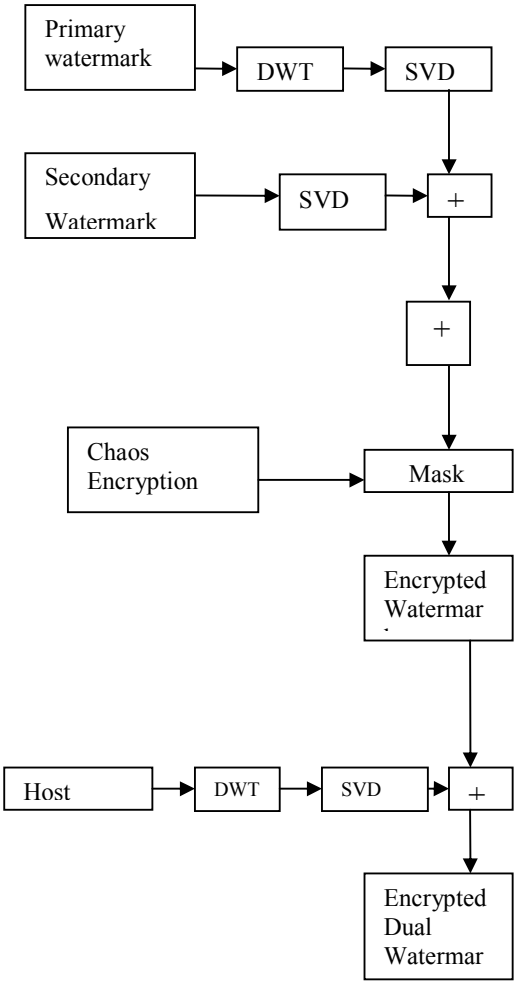


Fig. 1 Block Diagram of the proposed system

III. MODULES

In the proposed system, there are four modules, they are as follows:

1.Embedding secondary watermark into primary.
2.Encryption of watermarked primary image and embedding in the host image.
3.Attacks
4.Extraction of primary and secondary watermark from the host image.

A. Embedding secondary watermark into primary watermark

In two-dimensional DWT, each level of decomposition produces four bands of data denoted by LL, HL, LH, and HH. The LL subband can further be decomposed to obtain another level of decomposition. This process is continued until the desired number of levels determined by the application is reached. In DWT-SVD based watermarking, the singular values of the detail and approximate coefficients are extracted. The extracted singular values are modified to embed the watermark data.

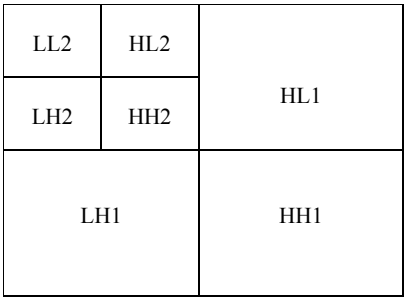


Fig. 2 DWT

Let A be a general real matrix of order m × n. The singular value decomposition (SVD) of A is the factorization:

$$A = U * S * V^T \quad (1)$$

where U and V are orthogonal(unitary) and S = diag(σ1, σ2, ..., σr), where σi, i = 1(1)r are the singular values of the matrix A with r = min(m, n) and satisfying :

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r \quad (2)$$

The first r columns of V the right singular vectors and the first r columns of U the left singular vectors.

Use of SVD in digital image processing has some advantages. First, the size of the matrices from SVD transformation is not fixed. It can be a square or a rectangle. Secondly, singular values in a digital image are less affected if general image processing is performed. Finally, singular values contain intrinsic algebraic image properties.

The singular values are resistant to the following types of geometric distortions:

i) Transpose: The singular value matrix A and its transpose A^T have the same non-zero singular values.
ii) Flip: A, row-flipped A_{rf}, and column-flipped A_{cf} have the same non-zero singular values.
iii) Rotation: A and A_r (A rotated by an arbitrary degree) have the same non-zero singular values.
iv) Scaling: B is a row-scaled version of A by repeating every row for L_1 times. For each non-zero singular value λ of A, B has $L_1\lambda$. C is a column-scaled version of A by repeating every column for L_2 times. For each nonzero singular value λ of A, C has L_2 λ. If D is row-scaled by L_1 times and column-scaled by L_2 times, for each non-zero singular value λ of A, D has L1L2λ.

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

v) Translation: A is expanded by adding rows and columns of black pixels. The resulting matrix A_e has the same non-zero singular values as A.

Algorithm – Embedding watermark.

i) Perform 1-level wavelet transform on the primary watermark.
Let us denote each sub-band with$W_{2\theta}$ where $\theta \varepsilon$ {LL, LH, HL, HH} represents the orientation.
ii) Perform SVD transform on secondary watermark,

$$W2 = U_{W2} * S_{W2} * V^{T}_{W2}.$$

iii) Perform SVD transform on approximation and all the detail parts of the primary watermark,

$$W1_{\theta} = U_{W1\theta} * S_{W1\theta} * V^{T}_{W1\theta}$$

where $\theta \varepsilon$ { LL, LH, HL, HH }.
iv)Modify the singular values of approximation and all the detail parts with the singular values of the secondary watermark as

$$S^*_{\theta} = \alpha S_{W2} + S_{W1\theta}$$

v) Obtain modified approximation and all the detail parts as

$$W^*_{1\theta} = U_{W1\theta} * S^*_{W1\theta} * V^{T}_{W1\theta}$$

where $\theta \varepsilon$ { LL, LH,HL, HH }.
vi) Perform 1-level inverse discrete wavelet transform to get the watermarked primary watermark.

B. Encryption

Chaos theory is a branch of mathematics which studies the behavior of certain dynamical systems that may be highly sensitive to initial conditions. This sensitivity is popularly referred to as the butterfly effect. As a result of this sensitivity, which manifests itself as an exponential growth of error, the behavior of chaotic systems appears to be random. That is, tiny differences in the starting state of the system can lead to enormous differences in the final state of the system even over fairly small timescales. This gives the impression that the system is behaving randomly.

Chaos-based image encryption techniques are very useful for protecting the contents of digital images and videos. They use traditional block cipher principles known as chaos confusion, pixel diffusion and number of rounds. The complex structure of the traditional block ciphers makes them unsuitable for real-time encryption of digital images and videos. Real-time applications require fast algorithms with acceptable security strengths. The chaotic maps have many fundamental properties such as ergodicity, mixing property and sensitivity to initial condition/system parameter and which can be considered analogous to some cryptographic properties of ideal ciphers such as confusion, diffusion, balance property.

A chaos-based image encryption system based on logistic map, in the framework of stream cipher architecture, is proposed. This provides an efficient and secure way for image encryption and transmission.

Logistic Map

One of the chaos functions that have been studied recently for cryptography applications is the logistic map. The logistic map function is expressed as:

$$X_{n+1} = rXn\ (1-X_n)$$

where X_n takes values in the interval [0,1]. It is one of the models that present chaotic behavior. The parameter *r* can be divided into three segments.
When X_0=0.3 and $r \varepsilon$ [0,3] the system works normal without any chaotic behavior. When $r \varepsilon$ [3, 3.57], the system appears periodicity. While $r \varepsilon$ [3.57, 4], it becomes a chaotic system with no periodicity.
We can draw the following conclusions:

i) When $r \varepsilon$ [0, 3.57], the points concentrate on several values and could not be used for image cryptosystem.

ii) For $r \varepsilon$ [3.57, 4], the logistic map exhibits chaotic behavior, and hence the property of sensitive dependence. So it can be used for image cryptosystem.

Algorithm – Image Encryption

i) The watermarked primary image is converted to a binary data stream.
ii) A random keystream is generated by the chaos-based pseudo-random keystream generator (PRKG).
iii) PRKG is governed by a logistic map, which is depended on the values of b, x_0.
iv)Through iterations, the first logistic map generates a hash value x_{i+1}, which is highly dependent on the input (b, x_0), is obtained and used to determine the system parameters of the second logistic map.
v) The real number x_{i+1} is converted to its binary representation x_{i+1}, suppose that L=16, thus x_{i+1} is {b1, b2, b3, … b16}. By defining three variables whose binary representation is X_l=b1…b8, X_h=b9…b16, we obtain X_{i+1}'=$X_l \oplus X_h$.
vi) Mask the watermarked primary image with the chaos values.
The generator system can be briefly expressed using the following logistic maps:

$$x_{i+1}=bx_i(1-x_i) \quad (1)$$

$$x_{i+1}' = X_{i+1}' = X_l \oplus X_h \quad (2)$$

$$WI'=WI_i \oplus X_{i+1}' \quad (3)$$

The encrypted watermarked primary image is then embedded into the host image and transmitted.

C. Attacks

To investigate the robustness of the algorithm, the watermarked image is attacked by Average and Mean Filtering, JPEG and JPEG2000 compression, Gaussian noise addition, Resize, Rotation and Cropping. After these attacks on the watermarked image, we compare the extracted watermarks with the original one. The watermarked

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

image quality is measured using PSNR (Peak Signal to Noise Ratio).

D. Extraction of watermarks

Decryption is the reverse iteration of encryption. After decryption of the watermarked primary image, the extraction process takes place.

Extracting Primary Watermark

The extraction technique for primary watermark is given asvfollows:

i)Perform 1-level wavelet transform on the host and the watermarked image. Denote each sub-band with W_θ and $\hat{W}_\theta$ for host and watermarked image respectively where $\theta \varepsilon$ {LL, LH, HL, HH} represents the orientation.

ii) The detail and approximation sub-images of the host as well as watermarked image is segmented into non overlapping rectangles.

iii) Perform SVD transform on all non overlapping rectangles of both images.

$$W_{1\theta} = U_{W\theta} * S_{W\theta} * V^{T}_{W\theta} \text{ and}$$

$$\hat{W}_\theta = U_{\hat{w}\theta} * S_{\hat{w}\theta} * V^{T}_{\hat{w}\theta}$$

where $\theta \varepsilon$ { LL, LH, HL, HH }.

iv)Extract singular values from primary watermark from all non overlapping rectangles as

$$S = S_{\hat{w}\theta} - S_{W\theta}/\beta$$

v) The primary watermark can be obtained as

$$W1' = U_{W1} * S * V^{T}_{W1}$$

Extracting Secondary Watermark

i)Perform 1-level wavelet transform on the primary watermark and the detected watermark W1'. Denote each sub-band with $W'_{1\theta}$ and $W_{1\theta}$ where $\theta \varepsilon$ {LL, LH, HL, HH} represents the orientation.

ii)Perform SVD transform on all subbands.

iii)The singular values of secondary watermark can be extracted as

$$S' = S'\ W'_{1\theta} - SW_{1\theta}/\alpha$$

iv) The secondary watermark can be obtained as

$$W2 = U_{W2} * S' * V^{T}_{W2}$$

IV.RESULTS AND DISCUSSION

The proposed algorithm is demonstrated using MATLAB. We have taken 8-bit gray scale tree image as host image of size 256 x 256 and for primary and
secondary watermark, we have used 8-bit gray scale lena image and boy image of sizes 128 × 128 and 64 × 64 respectively. The secondary watermark is embedded into primary and the watermarked primary is encrypted. For encryption, chaos encryption technique is used.

For embedding the encrypted watermarked primary into the host image, we have used 2-level of decomposition using Daubechies filter bank. For extracting both the watermarks, decryption is done using the chaos technique. The decrypted image is then used to extract the primary watermark and this is used for extracting the secondary watermark. In figures 2 and 3 all original, watermarked images and extracted watermarks are shown.

To investigate the robustness of the algorithm, the watermarked image is attacked by Average and Median Filtering, Gaussian noise addition,Resize and Rotation . After these attacks on the watermarked image, we compare the extracted watermarks with the original one.

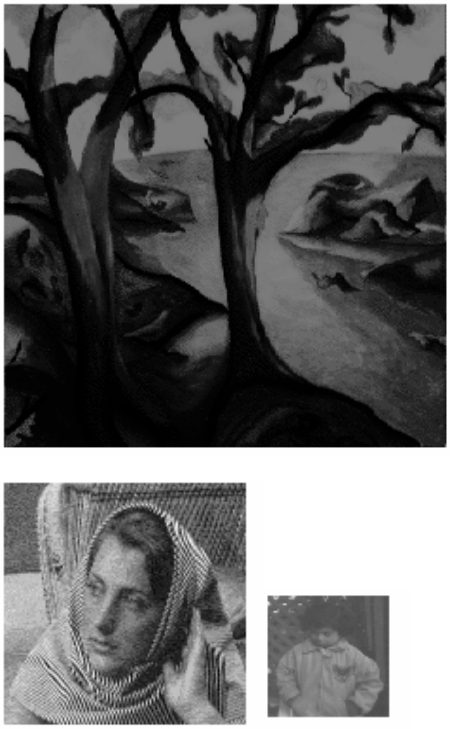


Fig. 2. Original Images a) Host image b) Primary watermark c) Secondary watermark

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

Fig. 3. Watermarked and extracted watermark images a) Watermarked Host image b) Watermarked Primary watermark c) Extracted Secondary watermark d) Extracted Primary watermark

Fig. 4. Median filtering Attack a) Attacked Host image b) Extracted Primary watermark c) Extracted Secondary watermark

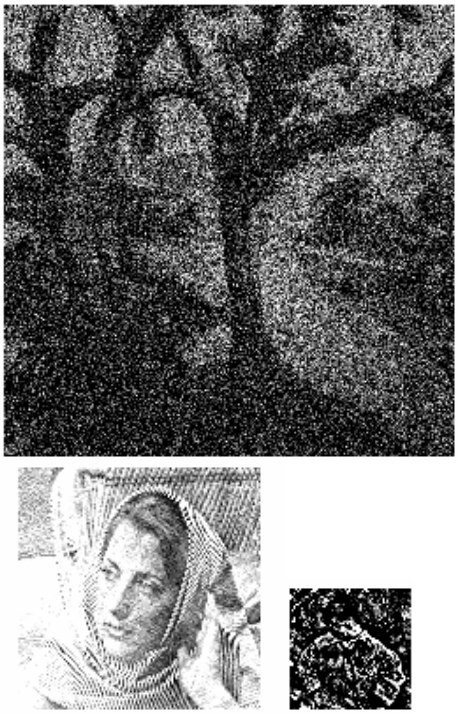


Fig. 5. Additive Gaussian Noise Attack a) Attacked Host image b) Extracted Primary watermark c) Extracted Secondary watermark

Fig. 6. Resize Attack (512 X 512) a) Attacked Host image b) Extracted Primary watermark c) Extracted Secondary watermark

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

Fig. 7. Rotation Attack (80^o Rotation) a) Attacked Host image b) Extracted Primary watermark c) Extracted Secondary watermark

Table 1. Correlation Coefficient of Extracted Primary and Secondary Watermark

Attacks	Primary	Secondary
Median Filtering	0.8967	0.4157
Additive Gaussian	0.8966	0.4161
Resize	0.8968	0.4154
Rotation	0.8969	0.4153

V.CONCLUSION

This paper deals with a novel dual watermarking scheme, which includes encryption, to improve rightful ownership, protection and robustness. An image encryption algorithm based on logistic map is proposed. A well-designed chaos-based stream cipher can be a good candidate and may even outperform the block cipher, on speed and security. In this, the key stream generator is based on coupled chaotic logistic maps that one logistic chaotic system generates the random changing parameter to control the parameter of the other. The watermarked primary image is encrypted using the chaos based encryption technique. Later it is embedded in the cover image and transmitted. The chaotic encryption scheme supplies us with a wide key space, high key sensitivity, and the cipher can resist brute force attack and statistical analysis. It is safe and can meet the need of image encryption.

For the extraction of watermark, a reliable watermark decryption scheme and an extraction scheme is constructed for both primary and secondary watermark. Robustness of this method is carried out by variety of attacks.

V.REFERENCES

[1] Gaurav Bhatnagar, Balasubramanian Raman and K. Swaminathan, " DWT-SVD based Dual Watermarking Scheme", IEEE International Conference on the Applications of Digital Information and Web Technologies (ICADIWT-2008), pp. 526-531.

[2]R. Liu and T. Tan, "An SVD-Based Watermarking Scheme for Protecting Rightful Ownership," IEEE Transactions on Multimedia, vol. 4, no. 1, 2002, pp. 121-128.

[3] E. Ganic and A. M. Eskicioglu, "Robust Embedding of Visual Watermarks Using DWT-SVD," Journal of Electronic Imaging, vol. 14, no. 4, 2005.

[4] Shubo Liu, Jing Sun, Zhengquan Xu, Jin Liu, "Analysis on an Image Encryption Algorithm", 2008 International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing, pp. 803- 806.

[5] Hossam El-din H. Ahmed, Hamdy M. Kalash, and Osama S. Farag Allah, "An Efficient Chaos-Based Feedback Stream Cipher (ECBFSC) for Image Encryption and Decryption", Informatica (2007) 121–129.

[6]Xiao-jun Tong, Ming-gen Cui, "A New Chaos Encryption Algorithm Based on Parameter Randomly Changing", IFIP International Conference on Network and Parallel Computing (2007) 303-307.

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

Effort minimization in UI development by reusing existing DGML based UI design for qualitative software development

P. K. Suri
Professor, Department of Computer Sc. & Application
Kurukshetra University, Kurukshetra,
Haryana, India

Gurdev Singh
Department of Computer Sc. & Application
Kurukshetra University, Kurukshetra,
Haryana, India

Abstract— **This paper addresses the methodology for achieving the user interface design reusability of a qualitative software system and effort minimization by applying the inference on the stored design documents. The pictorial design documents are stored in a special format in the form of keyword text [DGML tag based design]. The design document storage mechanism will expose the keywords per design stored. This methodology is having an inference engine. Inference mechanism search for the requirements and find the match for them in the available design repository. A match found will success in reusing it after checking the quality parameters of the found design module in the result set. DGML notations produces qualitative designs which helps in minimizing the efforts of software development life cycle.**

Keywords- User interface design, Reusable user interface, User interface design engineering, Qualitative user interface design repre-sentation, Qualitative design, Design inference, Effort minimi-zation in UI development.

I. INTRODUCTION

A new industry practice emerged in quality software development which deals with not only developing a run away solution for a problem but to get a better solution which could be reused. Generally, we spend a lot of our time and efforts in developing solutions for difficult, time consuming and mission critical problems. One of these solution set components is the UI designs document [1][2]. Design process takes maximum time and efforts and is never reused in the future.

Reusability is achieved up to some level in coding practices like OOPs, Component based developments, Active-X, technology where a piece of written code is reused after passing few checks for non discloser of the blueprints of the solution [4].

We propose that the root of reusability is there at the starting and bridging activities of the software development process that is design phase. This includes functional and user interaction design. We can consider the reusability at this level, such as UI design reusability [6][14]. If a UI design is made reusable, we can achieve the reusability in the development phases because this coding phase is very much driven from the software design.

This paper proposes a new approach in the context of the User interface design reusability. Reusable UI design methodology is of special concern with the effort minimization, which could be achieved by following our approach. This paper also discusses the notation of the design document storage for reusability.

Reusable UI design approach by inference is having the full impact of the strategic fitness in achieving the large and tough goals in small time and smartly. This approach will fit as getting a good domain solution in minimum efforts. A good review mechanism can also be imposed on the stored reusable UI design for assigning quality attributes.

Proposed approach eliminates the need for fresh efforts for UI design every time a solution is required. The UI design document having diagrams, images will be stored in textual format and every design stored will have a few keyword and properties [3][5]. Keyword submission per design facilitates the search against the requirement specifications and outcome of the result having the attribute tag of quality benchmark, like number of times the design is reused. The research also includes one tool, which helps in maintaining the central repository of the different design documents and inference mechanism on the stored design repository with in the organization.

II. DGML BASED UI DESIGN REPRESENTATION

The reusability of user interface design for a software system could be achieved by storing the graphical design documents in the form of text. The graphics-symbols of the design document can be represented in the form of text. Tools are available which allow the user to draw the user interface with diagrams. These tools however lack one aspect to store the design elements in the form of structured text as backend.

Our approach is to device some mechanism using which we can design the diagrams for user interface and represent them in the text format. So that when user creates one UI design and save it, it generates one text file also. The approach that we choose is to represent the text per design element is as a XML tag. So every design element will have a unique tag.

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

Initially the process start with making a new UI design document, which is XML, based. The system to create the UI design is having the collection of UI elements. All finite design elements have assigned names. User has to select and place these design elements while creating design document. When one UI design module gets completed, there will be one complete XML document associated with it. These UI design repository will be a collection of associated XML document per design as shown in following figure2.

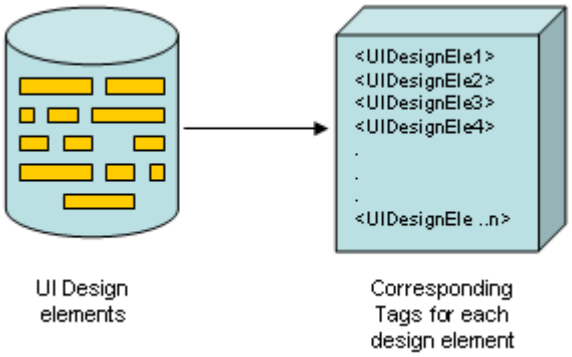


Figure 1: Every UI design element is having UI design element tag associated with it.

There will be a repository of designs after following this approach and every design is well defined in terms of XML tags. The text format of storing the design is having the predefined tags. Text format will be stored in XML notations. We say this as Design Markup Language and are having the tags for all knows design elements. When user picks one design element, its respective tag will be placed in its corresponding text file. When UI design complete for some scenario by following this approach, system will be having a complete representation in textual format in the form of tags.

The XML text file having the details of design figures is the basis of our research. The approach has many benefits. When pictorial representations are saved in text form, lots of possibilities are there which increase the overall development process [10][12]. Reusability of design search with in the available UI design, automatic generation of design skeleton on the basis of requirement document and available design text keywords, auto generation of test case scenarios.

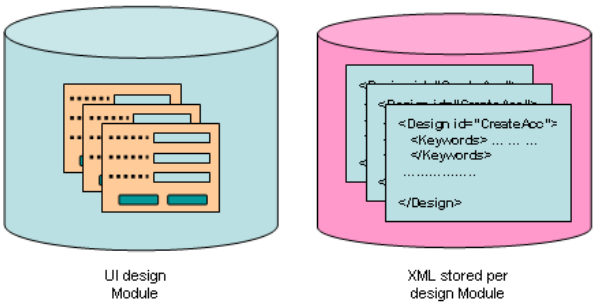


Figure 2: Each design module is having associated DGML document.

Along with the design element tags, the XML-DGML file is having the information with the help of special tags (like <keywords>), which is important and is a key for further reusability criteria implementation. This additional information is the list of keywords that designer wants to assign to his design. Each UI design is having one name, keywords and some attributes. Name identifies the design module, keywords are the handles for reusability, and attributes are factors, which design document gain after reusability. Experts review the UI design and assign scores. A good score by expert make design module a good candidate for reusability.

When some user wants to create the new design, he has to submit the UI requirement specification. The requirement specification document is analyzed and some keywords are evaluated from this. These are requirement specification keywords. Next step is that system looks into the available keywords of design module repository again these keywords. This special search process to find out required design from existing design is the design inference

The design repository will be stored for each design element in a centralized database of the organization. The design stored is of the atomic nature by representation of text format. Each elementary design element will be having one name and some associated keywords. These keywords will be reflected in the centralized repository.

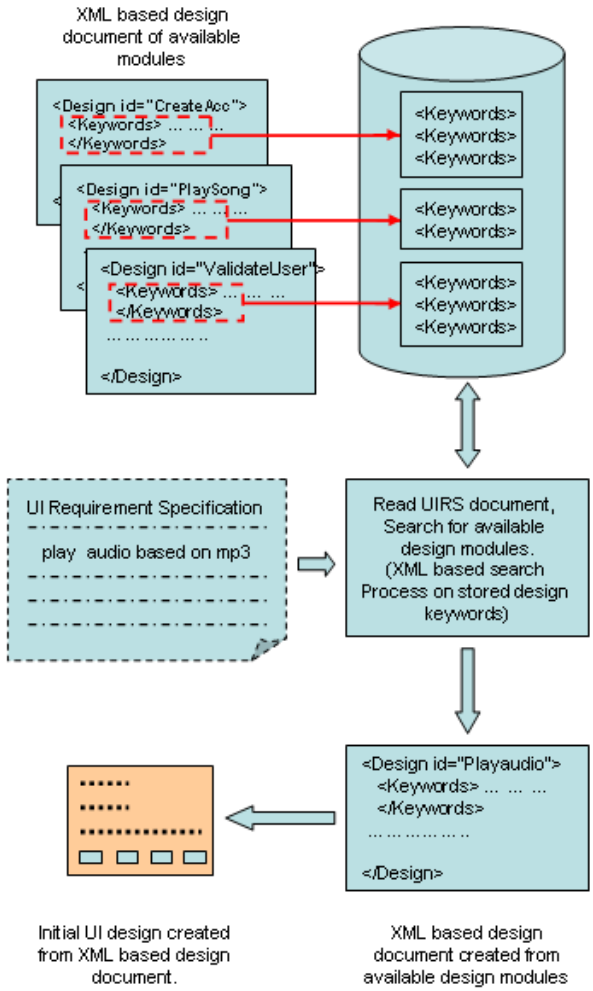


Figure 2: Mechanism for UI design reusability.

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

In software development life cycle, when analysis phase is completed and the requirements are defined, next stage is to move to design process. The existing design repository will help out in reusing the design elements.

Above figure shows the mechanism for UI design reusability. After following the approach for creating XML based UI design documents (DDML), we will be having a database of XML sheets for each design module. Each design module is having the name and keywords associated with it.

*< !ELEMENT name *>*

*< !ELEMENT keywords *>*

Inference will be preformed on the centralized design repository against the requirement specification keywords from the UIRS (User interface requirement specifications). This will be a type of look-in search process in the design repository keywords of elementary design. When search gets completed, it will generate a minimum UI design document having the maximum reusability of existing modules. This DGML document is arranged in order to get the maximum requirement fulfillment for required UI design.

Also, the result of the inference on stored design repository will be having the elementary design modules having attributes also. These attributes specify the design reusability factor for a module (DRF). More the reusability factor of a module indicates that module is strong candidate for reusability in new design. Reusing a UI design module will increase the reusability count factor by one.

III. WORKING

A.) Create DGML based UI design and generate design repository per design module.

Figure 4 is the logical representation while creation of new design. If user creates the new design and it is driven from some existing design element Dg..n, it will increment the design reusable factor DRF for Dg..n by one. This increment in DRF make makes design Dg..n a stronger candidate for further reusability

B.) Retrieve DGML based design repository for reusability and create initial proposed design layout.

The following flow chart (Fig 4) shows how the DRF (Design Reusable Factor) helps in reusing the existing UI design. First the search for the keywords in existing design will be performed. The outcome then will be sorted in descending order on the bases in DRF. The top value of result outcome shows that first one in the search is the best candidate for reusability

IV. EXPERIMENTAL DETAILS

For the experimental validation of the proposed approach of UI design reusability and effort minimization, we have considered five small projects having very small UI requirements. These projects are having maximum of six different modules. UI design requirement for these modules have been studied and we also consider one end user for our experiment to which we will give the finished design for review. (We assume and prepare this person with all the knowledge of the UI design, like its functionality scenario etc so that he may evaluate the UI design).

From requirement discussion to first draft, we record the time spend for each project. We also record the time consumed by the UI design architect. When this gets finished, we need to discuss it with the end user. Time record for the end user involvement was also recorded. We include the end user in the design phase for layout, look and feel type observation. We consider this as design prototype.

Table1 is the recorded of the efforts made by the UI design architect and end user for verification of the initial draft.

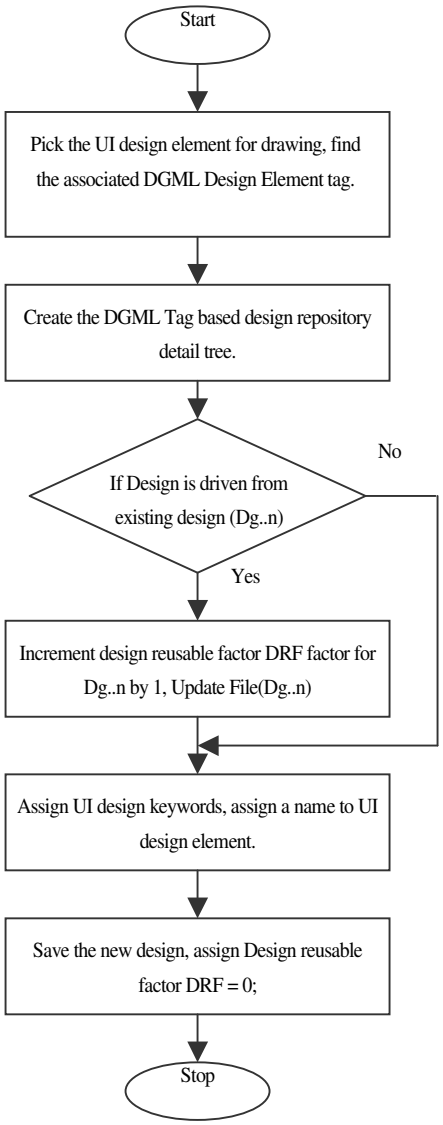


Figure 4: Create DGML based UI design

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

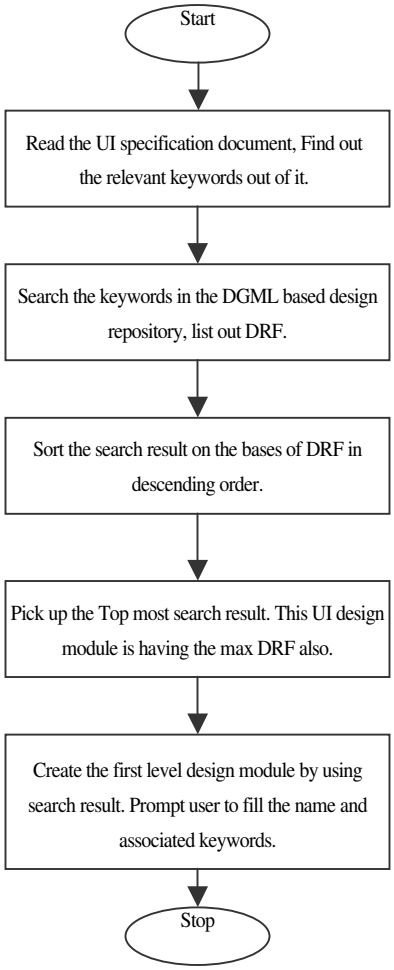


Figure 5: Retrieve DGML based design repository for UI design reusability

Table 1: Efforts (in Hrs) in UI design using conventional design methods.

Modules	Efforts in Hrs	
	Conventional UI Design efforts	End User involvement
Project1-4Modules	13	6
Project2-4Modules	29	8
Project3-5Modules	17	7
Project4-5Modules	8	4
Project5-5Modules	19	7
Project6-6Modules	20	7

This is the observation of the projects using conventional system and we will compare it with efforts consumed using our approach in next section with Table 2.

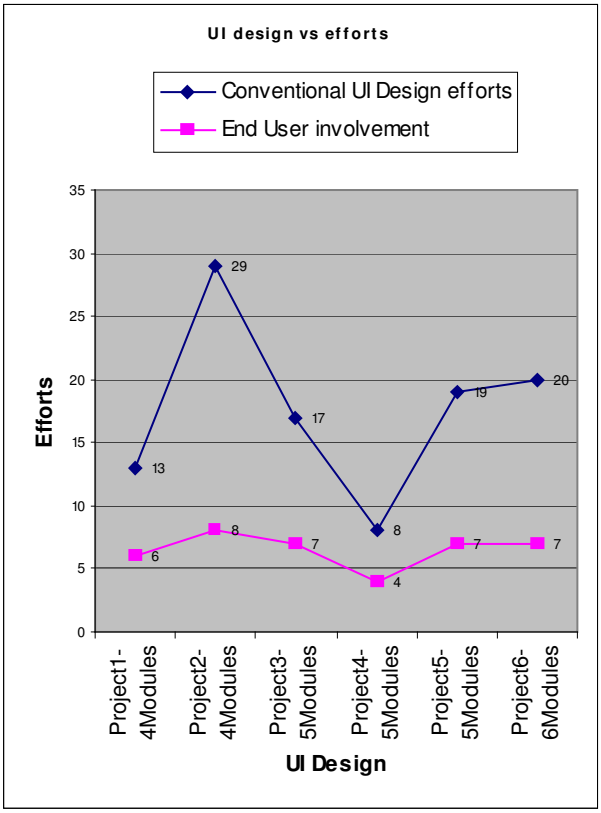


Figure 6: Efforts (in Hrs) in UI design using conventional design methods (Source: Table1).

The graph (Figure 6) represents that for a project of 4 modules, minimum time for design is 13 hrs where as maximum could be 29 hrs for a project of 4 modules. This variation shows that module count is immaterial but the important is nature of project. This observation when compared with the DGML based approach will shows very interesting results.

For our experiment with DGML based UI design, we are having 10 designs in the DGML storage and these are having 50% resemblance with the domain of the requirement. These design documents are having the keywords, which could be used during the search process.

The DGML based design experiment for the reusability of existing design requires the framing of the requirement specification in some special format. So, we have made the requirement specification document. This is having the keyword-based text instead of the plain English based scenario discussions. Same projects with special formatting of the requirement specification are submitted to DGML inference engine. This take very less time in identifying the candidate from available stored design and displays the results. The time taken in framing the requirement and getting the first level design from existing designs is shown in the following table (Table 2). Its approximately 1hr because this is not simply selection process, user have to spend some time (in few min) to

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

identify which one is appropriate on the bases of some ranks, like reusability factor.

Table 2: Efforts (in Hrs) in UI design using DGML based UI design methods.

Modules	Efforts in Hrs			
	Efforts in Framing Requirements	Effort in generating first level design	Agile client efforts	Total efforts in automation
Project1-4Modules	7	1	1	9
Project2-4Modules	10	1	2	13
Project3-5Modules	7	1	2	10
Project4-5Modules	3	1	1	5
Project5-5Modules	5	1	3	9
Project6-6Modules	7	1	2	10

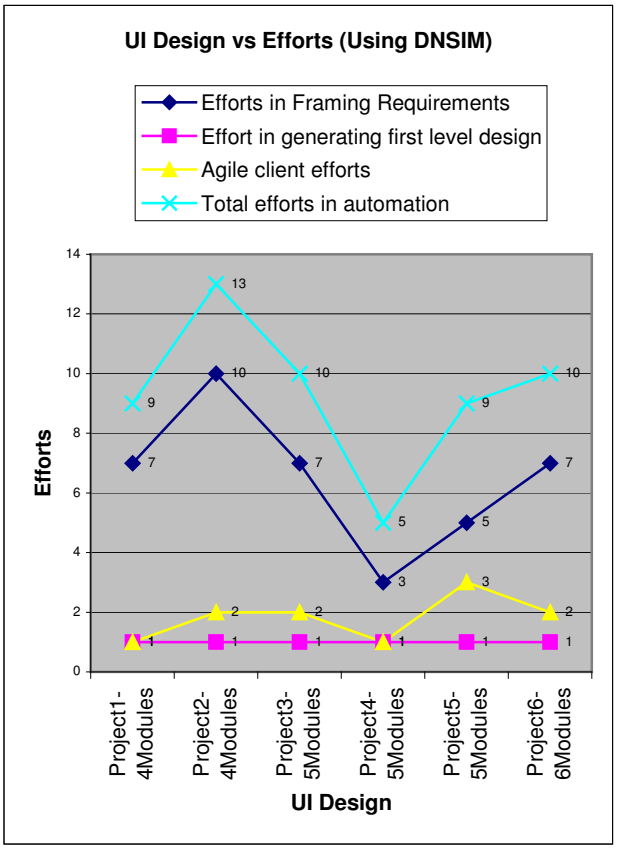


Figure 7: Efforts (in Hrs) in UI design using conventional design methods (Source: Table2).

Observations in Table2 are showing the improvement in the efforts consumed. The only efforts consumed are in the framing of the requirement in the special format.

The agile client efforts, which can be a part of team also need to spend small time as compared to time spend in table1 by the user [15][16]. We propose the involvement of the user in the design process with our mechanism. User interactions could be of the type of selecting one design layout out of available design layout produced by the system.

The total time consumed in the finalization which is summation of time consumed in framing requirement, design generation by system and client time to select. We represent this as TED, i.e. total effort required for DGML based design.

$$TED = \sum_{i \in SP} RFi \oplus DGi \oplus ACEi$$

Here RF is the efforts required for requirement framing for software project SP, DG is efforts in design generation and ACE is the agile client efforts. Figure 6 shows the graphical representation for the same.

As an observation from figure7 we can see the total efforts consumed in the proposed design are less than that of conventional system which is showing the effort minimization after using the DGML based design approach and then applying the DNSIM (Design notation storage and inference mechanism) for reusing the same for achieving the qualitative designs.

Table 3: Total Efforts comparison in using two approaches

Modules	Total efforts in DGML based approach	Total efforts in conventional UI design.
Project1-4Modules	9	19
Project2-4Modules	13	37
Project3-5Modules	10	24
Project4-5Modules	5	12
Project5-5Modules	9	26
Project6-6Modules	10	27

Graphs below (figure 8) shows the comparison and figures in the total effort minimization. DGML based design approach

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

uses requires the less efforts as compared to other methods. DGML based approach produces qualitative design, which helps in minimizing overall efforts in software development life cycle.

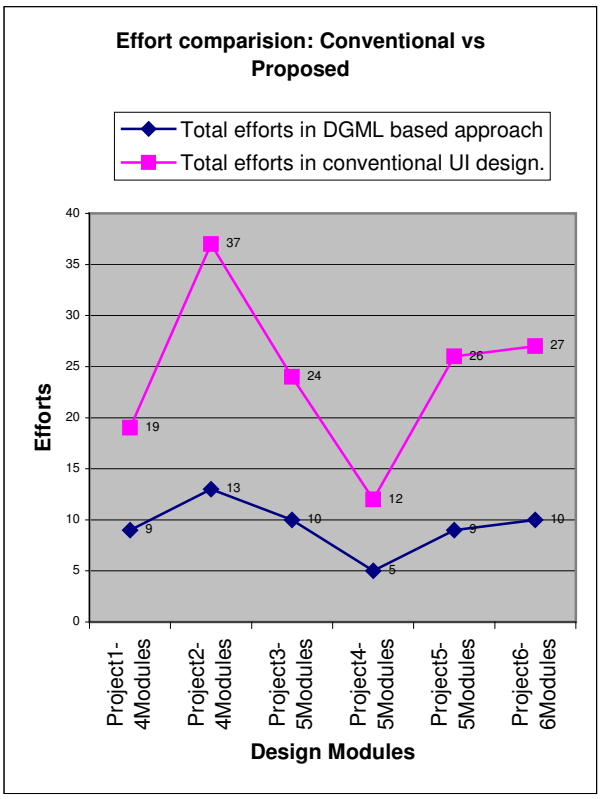


Figure 8: Total Efforts comparison: DGML based approach and conventional approach (Source: Table3).

Table 4: User involvement efforts comparison in using two approaches.

Modules	End User involvement. (Conventional)	Agile client efforts
Project1-4Modules	6	1
Project2-4Modules	8	2
Project3-5Modules	7	2
Project4-5Modules	4	1
Project5-5Modules	7	3
Project6-6Modules	7	2

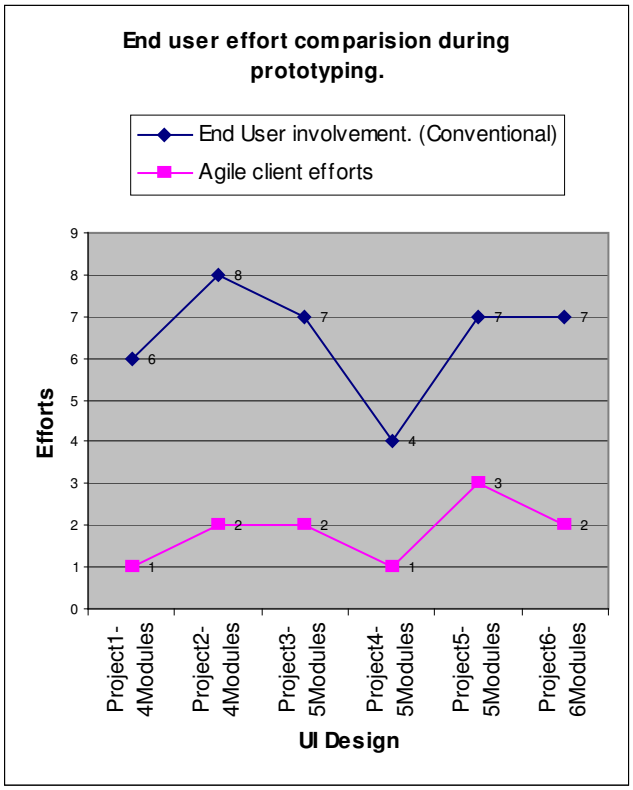


Figure 9: Total Efforts comparison: DGML based approach and conventional approach (Source: Table3).

The end user involvement and efforts for feedback are comparatively less in case of the DGML based approach. For same projects taken in our experiment, we found that there is a big difference in the end user involvement. As for project1 having 4 modules, the efforts consumed by the agile client for feedback are 6hrs for conventional approach where as its only 1hr required for the DGML based user interface design. This also represents that it's a qualitative design approach towards effort minimization and making SDLC less complex.

V. DISCUSSION

DGML based designs representation is a new approach for storing the design elements in the form of specially design DGML-XML tags. These are plain text files, so we can write programs to manipulate and analyze these files for further enhancement.

To simulate the effort minimization, we carry out above-mentioned experiment. It shows that the there is tremendous decrease in the user involvement. We carry out our experiment on five different projects having different number of modules. Table1 is showing the efforts in man hrs and end user involvement using the conventional, approaches of software design engineering. The graph in the figure6 depicts that there is big involvement of the end user in the design phase (we

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

consider agile methodology where customer is a member of development team) as for project2 29 hr efforts are from designer team and end user spend 8 hrs.

Table 2 shows the overall efforts for creating the first level design using DGML based approach. The total efforts in automation are less than that of the manual and conventional system. There is a significant minimization in the efforts of the agile client involvement as depicted from the figure7 and table2.

Figure8 shows the comparison between the total efforts of DGML based design and conventional design. In DGML based design, there is only on activity, which consumes time and that is the framing of the requirement specification document. This contains finding out the important keywords from the requirement specification document and using them later in the search process DGML based parsing.

Figure9 compares the involvement of the end user in the same projects using the two different approaches. It clearly shows that agile client involvement is reduced significantly after following the DGML based approach.

VI. CONCLUSION

The user interface design reusability leads to faster development of the application especially in the area where the UI is of prime importance. Big time and efforts, which are consumed in developing the core user interface design functionality, could be minimized. The user involvement in the design phase and efforts could be minimized in the design process. This is because the design elements are stored in the textual format and we can apply several kinds of search algorithms for finding the best solution for the problem. Further, design storage in specified format makes the system scalable and maintainable. The approach of the reusability of software design is novel. This paper is an initiative towards using DGML based UI design representation for reusability. The experiment carried on five projects for effort minimization using UI design reusability is showing interesting and valuable observation of the concept.

REFERENCE:

[1] Markopoulos, P., Wilson, S., Johnson and P., "Representation and use of task knowledge in a user interface design environment", Computers and Digital Techniques, IEEE Proceedings, Volume 141, Issue 2, March 1994, pp 79 – 84.

[2] Yam Li, "Intelligent User Interface Design Based on Agent Technology", software Engineering, 2009. WCSE '09. WRI World Congress, Volume 1, 19-21 May 2009, and pp 226 – 229.

[3] Maskers, J.; Layton, K.; Coning, K.; "Shortening user interface design iterations through real-time visualization of design actions on the target device", Visual Languages and Human-Centric Computing, 2009. VL/HCC 2009. IEEE Symposium on 20-24 Sept. 2009, pp 132 – 135.

[4] Hood Meier Halo; Afar, A.; "Tracing user interface design pre-requirement to generate interface design specification", Electrical Engineering and Informatics, 2009. ICEEI 'Aug. 2009, pp 287 – 292.

[5] Menthol Main; Sunhat, V.S.; No Sukaviriya; Ramachandra, T.;, "Using User Interface Design to Enhance Service Identification of Web Services", 2008. ICWS '08. IEEE International Conference on 23-26 Sept. 2008, pp 78 – 87.

[6] Tesarik, J.; Dolezal, L.; Kollmann, C.;,"User interface design practices in simple single page web applications", Applications of Digital Information and Web Technologies, 2008. ICADIWT 2008. First International Conference on the 4-6 Aug. 2008, pp 223 – 228.

[7] Bajwa, I.S.; Chaudhary, M.A.; "A Language Engineering System for Graphical User Interface Design (LESGUID): A Rule based Approach", Information and Communication Technologies, 2006. ICTTA '06. 2nd Volume 2, 0-0 0, pp 3582 – 3586.

[8] Ping Zhang; Small, R.V.; von Dran, G.M.; Barcellos, S.;, "Websites that satisfy users: a theoretical framework for Web user interface design and evaluation", System Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on Volume Track2, 5-8 Jan. 1999, pp 8

[9] Vila, J.; Beccue, B.; Furness, G.;, "User interface design for virtual reality: a research tool for tracking navigation", System Sciences, 1998., Proceedings of the Thirty-First Hawaii International Conference on Volume 6, 6-9 Jan. 1998, pp 464 – 472.

[10] Lindgaard, G.;, "Designing CSCW tools to support cooperative research", Computer Human Interaction Conference, 1998. Proceedings. 1998 Australasian, 30 Nov.-4 Dec. 1998, pp 61 – 62.

[11] Noble, J.; Constantine, L.L.;, "Interactive design metric visualization: visual metric support for user interface design", Computer-Human Interaction, 1996., Proceedings., Sixth Australian Conference on 24-27 Nov. 1996, pp 213 – 220.

[12] Balasubramanian, V.; Turoff, M.; "A systematic approach to user interface design for hypertext systems", System Sciences, 1995. Proceedings of the twenty-Eighth Hawaii International Conference on Volume 3, 3-6 Jan. 1995, pp 241 – 250.

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

[13] Burns, M.J.; Whitten, W.B., II;, "Innovation in the user interface design process", System Sciences, 1991. Proceedings of the Twenty-Fourth Annual Hawaii nternational Conference on Volume ii, 8-11 Jan. 1991, pp 70 – 78.

[14] Doane, S.M.; Lemke, A.C.;, "Using cognitive simulation to develop user interface design principles", System Sciences, 1990., Proceedings of the Twenty-Third Annual Hawaii International Conference on Volume ii, 2-5 Jan. 1990, pp 547-554.

[15] Stary, C.;, "User interface design: the WHO, the WHAT, and the HOW revisited", Computer Software and Applications Conference, 1995. COMPSAC 95. Proceedings., Nineteenth Annual International 9-11 Aug. 1995, pp 178-183.

[16] Quiroz, J.; Shankar, A.; Dascalu, S.M.; Louis, S.J.;, "Software Environment for Research on Evolving User Interface Designs", Software Engineering Advances, 2007. ICSEA 2007. International Conference on 25-31 Aug. 2007, pp 84 – 84.

[17] Lu Xudong; Wan Jiancheng;, "User Interface Design Model", Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Eighth ACIS International Conference on Volume 3, July 30 2007-Aug. 1 2007, pp 538-543.

AUTHOR INFORMATION

Dr. P.K. Suri received his Ph.D. degree from Faculty of Engineering, Kurukshetra University, Kurukshetra, India and master's degree from Indian Institute of Technology, Roorkee (formerly known as Roorkee University), India. He is working as Professor in the Department of Computer Science and Applications, Kurukshetra University, Kurukshetra–136119 (Haryana), India since Oct. 1993. He has earlier worked as Reader, Computer Sc. & Applications, at Bhopal University, Bhopal from 1985-90. He has supervised twelve Ph.D.'s in Computer Science and eleven students are working under his supervision. He has more than 125 publications in International/National Journals and Conferences. He is recipient of 'THE GEORGE OOMAN MEMORIAL PRIZE' for the year 1991-92 and a RESEARCH AWARD –"The Certificate of Merit – 2000"for the paper entitled ESMD – An Expert System for Medical Diagnosis from INSTITUTION OF ENGINEERS, INDIA. His teaching and research activities include Simulation and Modeling, Software Risk Management, Software Reliability, Software testing & Software Engineering processes, Temporal Databases, Ad hoc Networks, Grid Computing, and Biomechanics.

Gurdev Singh received his Masters degree in Computer Science from Department of Computer Science & Applications, Kurukshetra University, Kurukshetra, Haryana, India. Since 2002 he is working as Software Development Professional and had experience of working with MediaTek, and Siemens Information System, India. Currently he is working as senior software engineer for Samsung Electronics in Noida, India. He has completed projects in the field of software development for mobile devices. He loves to transfer user requirements in to piece of software. His interest includes work in the domain of software engineering, effort minimization in software development, qualitative software design and synchronization, software design representation methodologies and reusable software design techniques. He has written many papers in the related domain. He is fond of studying about the digital electronics and experimenting the same.

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

Medical Image Compression using Wavelet Decomposition for Prediction Method

S.M.Ramesh
Senior Lecturer, Dept. of ECE
Bannari Amman Institute of Technology
Erode, India
E-mail: smrameshme@yahoo.co.in

Dr.A.Shanmugam
Professor, Dept. of ECE
Bannari Amman Institute of Technology
Erode, India
E-mail: dras_bit@yahoo.com

Abstract— ***In this paper offers a simple and lossless compression method for compression of medical images. Method is based on wavelet decomposition of the medical images followed by the correlation analysis of coefficients. The correlation analyses are the basis of prediction equation for each sub band. Predictor variable selection is performed through coefficient graphic method to avoid multicollinearity problem and to achieve high prediction accuracy and compression rate. The method is applied on MRI and CT images. Results show that the proposed approach gives a high compression rate for MRI and CT images comparing with state of the art methods.***

Keywords- Correlation coefficient, Selection of predictor, Variable, DPCM, Arithmetic coding.

I. INTRODUCTION

Image compression is required to minimize the storage space and reduction of transmission cost. Medical images like MRI and CT are Special images require lossless compression as a minor loss can cause adverse effects. Prediction is one of the techniques to achieve high compression. It means to estimate current data from already known data [1].

The advance image compression techniques for medical images are JPEG 2000[2] which combines integer wavelet transform with Embedded Block Coding with Optimized Truncation (EBCOT). It is an compression rate. Context based adaptive compression rate. Context based adaptive advanced technique which provides high lossless image codec (CALIC) is offered by Wu and Memon [3]. They utilized the prediction in the original CALIC but offered inter band prediction technique for remotely sensed images. A better technique for lower quality ultrasound images is offered by Przelaskrwski [4] to achieve a high compression rate. Buccigrossi and Simoncelli [5] made a statistical model and used conditional probabilities for prediction. That is a lossy method called Embedded Predictive Wavelet Image Coder(EPWIC). Yao-Tien Chen & Din-Chang Tseng proposed the Wavelet-based Medical Image Compression with Adaptive Prediction (WCAP).They used correlation analysis of wavelet coefficients to identify the basis function and further for prediction. They used lifting integer wavelet scheme for image decomposition. It is a lossless scheme to achieve highest bit rate per pixel (bpp). In (WCAP) they used backward elimination method for predictor variable selection and quantized the prediction error into Three levels (-1, 0, 1) to achieve the higher compression rate. They used DPCM for coarse bands and finally used adaptive arithmetic coding.

To achieve a high compression rate for medical images we propose wavelet based compression scheme using prediction, "Medical Image Compression using wavelet decomposition for Prediction method". The scheme uses the correlation analysis of wavelet coefficients like WCAP but adds simplicity and accuracy by excluding the requirement of selection of basis function and quantization of prediction error in coarse bands. A simple, graphic method for variable selection is introduced. The proposed scheme block diagram shown in figure.1, consists of six major stages including, image decomposition, correlation analysis of wavelet coefficients, development of prediction equation for each sub band, predictor variable selection using graphic method, arithmetic coding and reconstruction of original image. The remaining sections of this paper is organized as following, section-2 covers the Lifting Wavelet Transform of group of similar images, predictor variable selection in section- 3, experiments, results and discussion in section 4 and conclusion in the final.

II. LIFTING WAVELET TRANSFORM OF A GROUP OF IMAGES AND CORRELATION ANALYSIS

The wavelet transform is a very useful technique for image analysis and Lifting Wavelet Transform is an advance form of wavelet transform which allows easy computation, better reconstruction of original image and close approximation of some data sets. The inter scale and intra scale dependencies of wavelet coefficients are exploited to find the predictor variable. All coefficients of current, parent and aunt sub bands of each processing coefficient are found.

The correlation coefficient is based on variance and co-variance. Covariance is always measured between two matrices or dimensions while the variance is measured for a dimension with itself. The formulae for variance and covariance are as following.

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

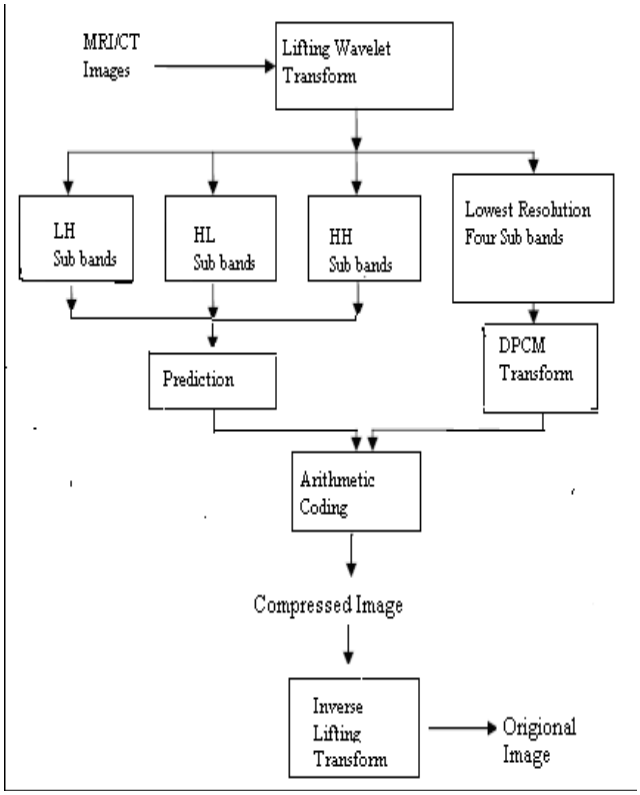


Figure1. Block diagram of proposed scheme

$$\mathbf{Var\ (X)} = \frac{\sum_{i-1}^{n} (X_i - X^b)}{n - 1} \quad (1)$$

$$\mathbf{Var\ (Y)} = \frac{\sum_{i-1}^{n} (Y_i - Y^b)}{n - 1} \quad (2)$$

$$\mathbf{Cov\ (X, Y)} = \frac{\sum_{i-1}^{n} (X_i - X^b)(Y_i - Y^b)}{n - 1} \quad (3)$$

$$\mathbf{R_{XY}} = \frac{Cov(X, Y)}{Var(X)\ Var(X)} \quad (4)$$

Where X^b & Y^b are means of X and Y and R is correlation coefficient.

III. PREDICTION AND PREDICTOR VARIABLE SELECTION

Our predictor is based on the linear prediction model containing k independent variables, can be written as [1]…

$$y = a_1x_1 + a_2x_2 + \text{---------} + a_kx_k \quad (5)$$

In this equation y is dependent variable and x1, x2…..xk are independent predictor variables. Where as a1, a2, …ak are predictor model parameters. To avoid the multicollinearity problem the number of predictor variables should be reduced. There are multiple methods to reduce the predictor variables. The best method is one which gives accurate prediction.

In the proposed method we use coefficient graphic method for selection of prediction variables. It is a simple method in which predicted and original coefficients of a sub bands are plotted for comparison. Different combination of variables is tested to select the combination which best matches the original sub band coefficient graph. This is a simple and easy method.

The sequence of prediction is from course sub band to fine sub band and from left up coefficient to the right down coefficient. The fine sub band coefficients are predicted from coarse sub band coefficients and coarse sub band coefficients are not predicted. The course sub band coefficients are than processed by Differential Pulse Code Modulation (DPCM), which is most common predictive quantization method. This method exploits correlation between successive samples of source signals and encoding based on the redundancy in sample values to give lower bit rate. This method encodes the prediction error between the sample value and its predicted value to give high compression ratio. The coarse and fine sub band coefficients are than arithmetically encoded.

IV. EXPERIMENTS, RESULTS AND DISCUSSION

Two MRI and two CT gray scale standard test images as shown in figure 2 of size128*128 have been taken from world wide web for experiments and comparisons. MATLAB 7.0 has been used for the implementation of the proposed approach and results have been conducted on Pentium-1V, 3.20 GHz processor with a memory of 512 MB. BPP (Bits Per Pixel) metric is evaluated to compile compression result. Every image was decomposed into three scales with 10 wavelet sub bands.

Eleven correlation coefficients to the dependent c are selected which are Parent, Parent-East, Parent-West, Parent-South, Parent-North, North, North-East, North-West, West, Aunt 1 and Aunt 2 . The prediction equations for coefficients for different sub bands are derived one by one. Using the coefficient graphic method, prediction variables are selected for each sub band to get accurate prediction. The compression rates for the 4 medical images using the proposed, "MICWDP" method with two famous lossless methods: SPHIT and JPEG2000 is shown in Table1. Due to proper selection of predictor Variables, proposed approach almost achieves the highest compression rates. The comparison of average encoding / decoding time of two lossless compression methods is also shown in Table 2. The proposed method

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

makes use of "coefficient graphic method", approach successfully on medical images to get the best results.

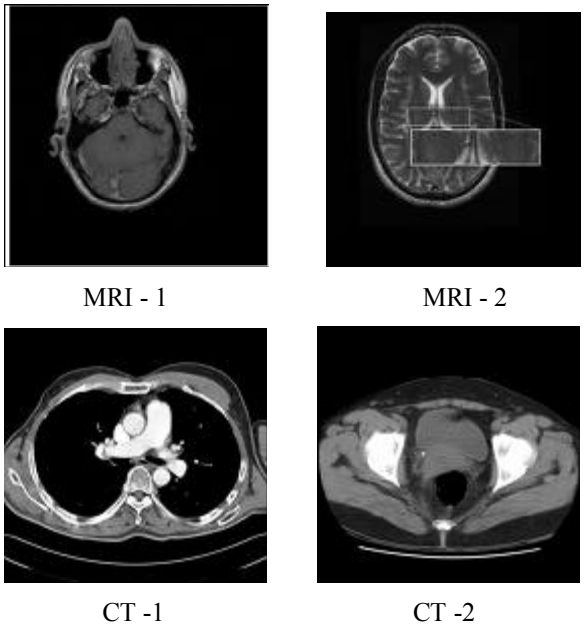


Figure. 2 2*MRI and 2*CT images taken for experiment

TABLE I
COMPARISON OF COMPRESSION RATE IN BITS/PIXEL OF DIFFERENT METHODS WITH PROPOSED METHOD

Type	Method		
	SPHIT	JPEG 2000	Proposed
MRI -1	2.53	2.42	1.45
MRI -2	3.11	3.12	1.51
MRI Average	2.82	2.77	1.48
CT -1	1.45	1.32	1.41
CT -2	1.79	1.82	1.43
CT Average	1.62	1.57	1.42

TABLE II
COMPARISON OF ENCODING/DECODING TIME OF DIFFERENT COMPRESSION METHOD

Type	Method		
	SPHIT	JPEG 2000	Proposed
MRI Average	1.7 / 1.9	0.8 / 0.8	2.54 / 3.13
CT Average	2.4 / 2.8	1.2 / 1.0	2.60 / 3.15
Average	2.05 / 2.35	2.00 / 0.9	2.57 / 3.14

V. CONCLUSIONS

In the proposed MICWDP approach, compression rate has been improved by exploiting dependencies among wavelet coefficients [1]. A new method, i.e coefficient Graphic Method is used to avoid ulticollinearity problem which is the main contribution of this method. Comparing with the SPHIT, JPEG2000 and proposed achieves the highest compression rate.

REFERENCES

[1] Yao-Tien Chen and Din-Chang Tseng, Wavelet-based medical Image compression with adaptive prediction. In: proc,International symposium on Intelligent Signal Processing and Communication Systems, December 2005-Hong Kong p.825-8 and Computerized medical Imaging and graphics 31(2007) 1-8

[2] Krishnan K.marcellin MW, Bilgin A, Nadar M.Prioritization of compressed data by tissue type using JPEG2000, In:proc.SPIE medical imaging 2005-PACS and imaging informatics.2005. p.181-9

[3] Wu X, Memon N. Context-based, adaptive, lossless Image coding. IEEE Trans Image Process 1997;6(5):656-64.

[4] Przelaskowski A. lossless encoding of medical images:hybrid modification of statistical modeling-based conception. J Electron Imaging 2001;10(4):966-76.

[5] Bussigrossi RW, Simoncelli EP. Image compression via joint Statistical characterization in the wavelet domain. IEEE Trans Image Process 1999;8(12);1688-701.

[6] Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, 2nd Edition, Printice Hall Inc, 2002.

[7] Ian Kaplan, Basic lifting scheme wavelets, February 2002(revised)

[8] Lindsay I Smith, A tutorial on principal component analysis, 26 February 2002.

[9] Majid Rabbani and Paul W. Jones, Image compression techniques.

[10] In H. Witten, Radford M. Neal and John G. Cleary, Arithmetic Coding for data compression.

[11] Mark Nelson, Arithmetic coding, Dr. Dobbs Journal,February, 1991.

AUTHORS PROFILE

Dr.A.Shanmugam received the B.E, degree in Electronics and Communication Engineering from PSG College of Technology., Coimbatore, Madras University, India in the year 1972 and the M.E, degree in Applied Electronics from College of Engineering, Guindy, Chennai, Madras University, India in the year 1978 and received the Ph.D. in Computer Networks from PSG College of Technology., Coimbatore, Bharathiyar University, India in the year 1994.From 1972 to 1976, he served as a Testing Engineer at Test and Development Center, Chennai, India. From 1978 to 1979, he served as a Lecturer in the Department of Electrical Engineering, Annamalai University, India. From 1979 to 2002, he served different level as a Lecturer, Asst.Professor, Professor and Head in the Department of Electronics and Communication Engineering of PSG College of Technology, Coimbatore, India. Since April 2004,he assumed charge as the Principal, Bannari Amman Institute of Technology, Sathyamangalam, Erode, India. He works in field of Optical Networks, broad band computer networks and wireless networks, Signal processing specializing particularly in inverse problems, sparse representations, and over-complete transforms.

Dr.A.Shanmugam received "Best Project Guide Award" five times from Tamil Nadu state Government. He is also the recipient of "Best Outstanding Fellow Corporate Member Award" by Institution of Engineers (IE),India - 2004 and "Jewel of India" Award by International Institute of Education and Management, New Delhi–2004 and "Bharatiya Vidya Bhavan National Award for Best Engineering College Principal 2005" by Indian Society for Technical Education (ISTE). "Education Excellence Award" by All India Business& Community Foundation, New Delhi.

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

Dr.A.Shanmugam

Mr.S.M.Ramesh

S.M.Ramesh received the B.E degree in Electronics and Communication Engineering from National Institute of Technology (Formerly Regional Engineering College), Trichy, Bharathidhasan University, India in the year 2001 and the M.E, degree in Applied Electronics from RVS College of Engineering and Technology, Dindugal, Anna University, India in the year 2004. From 2004 to 2005, he served as a Lecturer in the Department of Electronics and Communication Engineering, Maharaja Engineering College, Coimbatore, India. From 2005 to 2006, he served as a Lecturer in the Department of Electronics and Communication Engineering, Nandha Engineering College, Erode, India. Since June 2006, he served as Sr.Lecturer, in the Department of Electronics and Communication Engineering Bannari Amman Institute of Technology, Sathyamangalam, and Erode, India. He is currently pursuing the Ph.D. degree, working closely with Prof. Dr.A.Shanmugam and Prof Dr.R.Harikumar.

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

High Performance Hybrid Two Layer Router Architecture for FPGAs Using Network-On- Chip

P.Ezhumalai [1] Dr. C.Arun[2] S.Manojkumar[3] Dr.P.Sakthivel[4] Dr.D.Sridharan[5]
[1&3]Department of Computer Science and Engineering
Sri Venkateswara College of Engineering, Pennalur, Sriperumbudure-602105, Chennai, Tamilnadu, India
[1]
2Department of Electronics & Communication Engineering
Rajalakshmi Engineering College Thandalam, Chennai - 602 105, Tamilnadu, India
[3]
[4&5]Department of Electronics & Communication Engineering
College of Engineering, Guindy Anna University, Chennai, Tamilnadu

Abstract— **Networks-on-Chip is a recent solution paradigm adopted to increase the performance of Multi-core designs. The key idea is to interconnect various computation modules (IP cores) in a network fashion and transport packets simultaneously across them, thereby gaining performance. In addition to improving performance by having multiple packets in flight, NoCs also present a host of other advantages including scalability, power efficiency, and component re-use through modular design. This work focuses on design and development of high performance communication architectures for FPGAs using NoCs**

Once completely developed, the above methodology could be used to augment the current FPGA design flow for implementing multi-core SoC applications. We design and implement an NoC framework for FPGAs, Multi-Clock On-Chip Network for Reconfigurable Systems (MoCReS).We propose a novel micro-architecture for a hybrid two-layer router that supports both packet-switched communications, across its local and directional ports, as well as, time multiplexed circuit-switched communications among the multiple IP cores directly connected to it. Results from place and route VHDL models of the advanced router architecture show an average improvement of 20.4% in NoC bandwidth (maximum of 24% compared to a traditional NoC). We parameterize the hybrid router model over the number of ports, channel width and bRAM depth and develop a library of network components (MoClib Library). For your paper to be published in the conference proceedings, you must use this document as both an instruction set and as a template into which you can type your own text. If your paper does not conform to the required format, you will be asked to fix it.

Keywords: ***Core Based Design, FPGA, Network on Chip (NoC), On Chip Communication, MoCReS, System on Chip (SoC),***

I. INTRODUCTION

The two main concerns with NoC designs that are strictly packet-switched are the control and serialization overhead involved in transferring data between IP cores that are placed close to each other in the FPGA. In order to ensure high throughput between these cores, we advocate time-multiplexed circuit-switched connections. In addition to this mode of transfer, the router also preserves the online nature of communication between farther cores through the packet-switched layer. The area efficient MoCReS architecture is modified to support both the above mentioned layers of operation. The design goals and issues involved in the hybrid two-layer architecture are presented in this paper. We also develop a SystemC model of our router for both functionally verifying the design as well as to vary its specifications and obtain the performance results rapidly through simulation. We present the results and analysis of the novel router architecture in this paper.

We target our proposed NoC framework for reconfigurable computing platforms and therefore we restrict our discussions in this section primarily to existing FPGA based NoCs. NoCs were introduced into the FPGA domain mainly to simplify tile-based reconfiguration [1] [2], and its potential as effective communication architecture is largely unexplored [7].

Research in [5] [6] address the capabilities of FPGAs to support NoC based multi-processor applications. Hilton et al. [4] incorporate flexibility into their design for FPGA based circuit-switched NoCs. However, their strictly circuit-switched router suffers from signal integrity and path reservation issues which we overcome in our design. SoC BUS [8] proposes a circuit-switched router with a packet based setup. Here, control packets are responsible for setting up strict circuit-switched connections, which is different from our two-layer approach. Research in [9] [4] [3] also present FPGA based NoCs. The above designs ignore implementation level area-performance trade-offs while proposing the architecture, thereby limiting to a system-level performance analysis. To the best of our knowledge, this is the first work to propose an FPGA-suitable hybrid router architecture integrated with an automatic topology synthesis framework that

satisfies the bandwidth requirements of an application while optimizing its area overhead

II. Motivation

Packet-switching performs online scheduling by dynamically negotiating communication between the cores. An alternate technique, namely circuit-switching offers high through- put dedicated connections to overcome the performance drawbacks in packet-switching by scheduling time-multiplexed communication across the cores. Even though this static scheduling requires all the communication patterns to be known before hand, it can pro- vide a very high throughput with marginal area overhead (for storing schedules). We propose a modified router architecture which interfaces multiple IP cores to the router and supports packet-switching for inter router transfers and time-multiplexed circuit- switching for IP cores connected to the same router. This technique also eliminates the latency in req/grant protocol, serialization and control overheads for data transfers between cores placed close to each other in FPGAs and mapped to the same router.

A. Packetization and Control Overheads

In this section, we quantify the overheads associated with the existing baseline approach (MoCReS). Control and Packetization are the two main overheads associated with the MoCReS framework.

Control Overhead: In MoCReS, connections between various ports are established through a req/grant protocol which involves round-robin arbitration in the case of common ports requests (conflicts). We see that it takes at least 6 cycles for the data at the input port to appear at the output of a router (as input to the downstream router/local IP). This setup latency is a fixed overhead in addition to the delays due to network congestion.

Packetization Overhead: Due to the nature of interconnection network, the channel width between ports/routers are limited to a fixed size (8 bits in MoCReS, baseline version). Due to this fixed channel width, the communication data that is to be sent over the network must be quantized into flits. Variable number of flits constitutes a packet. If F is the number of flits in a packet and b is the channel width, then F /b is the serialization latency associated with the communication.

III. Architecture Description

In this section, we first present the modified router micro-architecture, followed by its architectural advantages and design issues involved. The network topologies along with the Flow controls for the packet-switched layer are kept the same as presented in this paper [10]. Network Topology: Mesh networks have minimum area overhead (reduced long lines) [5] [10], low power consumption and map well to the underlying routing structure of FPGAs. Hence, we choose a mesh topology to optimize logic and routing in FPGAs, and to provide sufficient resources for the IP cores.

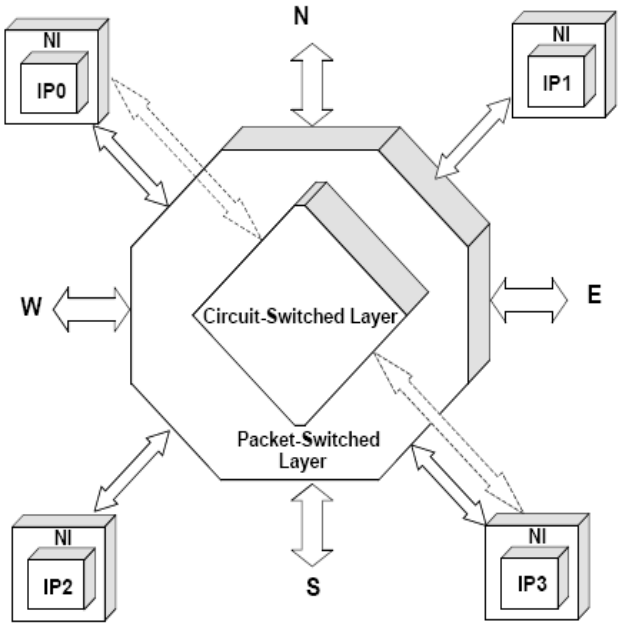


Figure 1 : Hybrid Two-Layer Router Architecture

Flow Control: Our router supports multi-clock virtual cut-through flow control with a deadlock-free XY routing. The switch complexity involved in the above choice is more suitable for a light-weight implementation [10].

A. Cross-Point Matrix

Architecture Modifications: The modified switch is comprised of two layers of operation: a high throughput time-multiplexed circuit-switched layer (C-layer) and a multi- clock packet-switched layer (P-layer). Variable number of IP cores connected to the switch participates in the C-layer, thereby achieving guaranteed throughput and more predictable latencies between IP cores placed close to each other in the FPGA.

Figure 1 presents the novel two-layer hybrid router architecture. This modified router has four local IP ports, in addition to the four directional ports. Further, in this case two of the four local IPs (IP 0, IP 3) are participating in the time-multiplexed circuit-switched layer. Using the packet-switched layer, all the four IPs can communicate to the neighboring routers through the

directional ports. The cross-point matrix is multiplexer based, as opposed to providing connections for each virtual channel. The following are the design issues involved with the cross-point.

Packet-Switched Cross-Point: In the packet-switched layer, the directional input ports (N, E, and S, W) are multiplexed to every local port. Therefore cross-point connections are introduced to support these additional local ports. However, all the connections between the local ports in this layer are removed, as they are connected in the circuit- switched layer. The ports connected through the C-Layer (IP 0, IP 3) cannot participate in the P-Layer to transfer data between themselves. This translates into gain in area which we utilize to increase the bandwidth available.

.

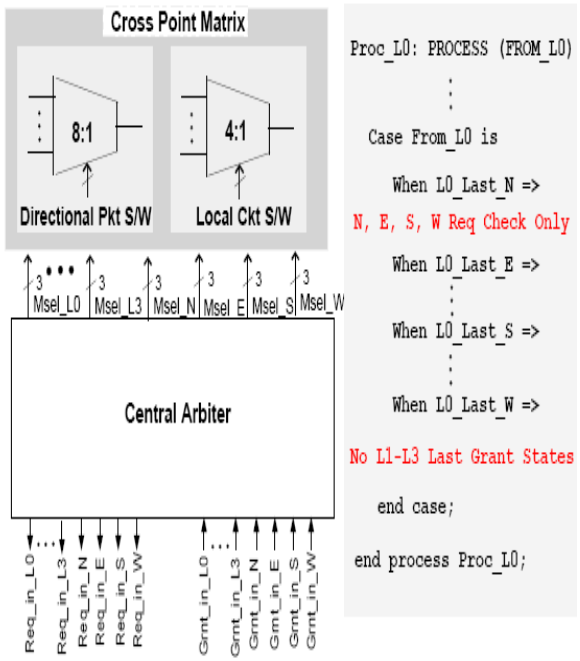


Figure 2: Modified Central Arbiter Model

Circuit-Switched Cross-Point: Let L_i is the total number of local IPs and P_i is the number of ports participating in the circuit-switched layer. The bus width of this cross-point is currently set to 32 bits in order to support a very high bandwidth. Further, this cross-point can handle a maximum of P_i high throughput parallel connections. The scheduling memory configures this cross-point during various time slots.

Router Channel Widths: Due to high throughput requirement between the cores participating in the circuit-switched layer, we set the channel width to 32 bits (corresponding to the data width of micro blaze soft processor). In the packet-switched layer, we retain the bus width of MoCReS (8 bits/channel). However, choice of an appropriate channel width is a trade off between resources available and bandwidth required.

B. Central Arbiter

The Central Arbiter is responsible for configuring the simultaneous connections by setting the cross-point in the P-Layer. We run parallel FSMs to ensure that no queuing takes place between requests. As long as the participating IPs request mutually exclusive ports, the connections happen parallel. In case of queuing/conflicts, the arbitration is performed through the round robin approach. The IPs that participates in the C-Layer will not need arbitration between them in the P-Layer. We perform state reduction in the FSMs corresponding to those inter-local port connections .i.e in correspondence with the inter local IP connections that are removed (Section 3.1) in the packet-switched layer. The Central Arbiter is also customized to not support states for these connections. The simplicity of round-robin arbitration coupled with the above state reduction translates into significant area savings. Figure 2 shows the modified central arbiter model.

C. NI Design

The network interface arbitrates the choice of packet/circuit switched layer and is also responsible for supporting variable size packets. Mode Switching: Upon receiving the target IP co-ordinates, it triggers the mode signal to decide if the packet will be decoded to leave the router or the cross point is triggered in circuit switch mode.

Variable Packet Sizes: In during packet-switched transfer, the network interface is also responsible for encoding the header with:

1. Packet Size (As a fraction of bRAM depth)

2. X co-ordinate of destination IP

3. Y co-ordinate of destination IP

The packets transferred through the network can be broadly classified as control (lesser number of flits) or data. Therefore, the packets will be of varied sizes. The NI encodes the packet size as a fraction of the total bRAM depth along with the header. This novelty improves buffer utilization, thereby increasing the performance of the NoC.

D. Design Parameters

In order to quickly explore the NoC design space, we have parameterized the structural

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

VHDL model of our router for:

1. Total number of ports
2. Channel width
3. Virtual Channels/port
4. Number of ports participating in the C-Layer

By varying the above parameters, we develop a component library, MoC lib which we use to characterize variants of the router for area and operating frequency.

IV. Architectural Advantages

Bandwidth Increase: Bandwidth available in a switch is the product of the number of ports, operating frequency and channel width. The C-layer has minimum logic overhead with no buffering and can operate at a clock rate significantly higher than the P-layer. Furthermore, increasing the number of ports also scales the available band- width in a switch. Moreover, the absence of control/serialization overheads (req/grant) also increases the throughput.

Power Savings: The amount of logic required for the NoC reduces with router count, thereby saving static power. Further, with increasing number of ports within a router, the average packet latency is also reduced [9]. Therefore dynamic power drops considerably with reduction of router hops.

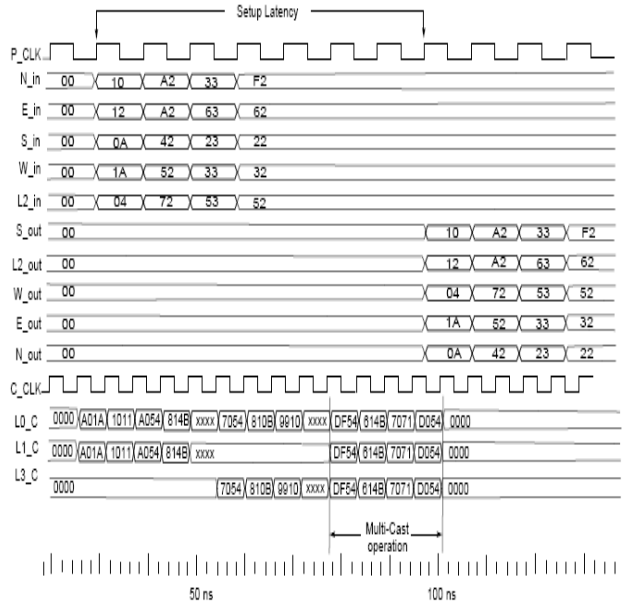


Figure 3: SystemC Simulation

Guaranteed Throughput: The time-multiplexed nature of the C-Layer scheduling provides good Quality of Service (QoS) to the application, particularly, between cores placed close to each other. Otherwise, the NoC would have to support area expensive QoS protocols to ensure the required bandwidth.

Inherent Multi-Cast Capability: The cross-point in the C-layer can be configured simultaneously for a multi-cast (one to many destinations) operation among IPs connected to the same router without any penalty in performance. Further, this capability also optimizes the area required for storing the schedules (with fewer bits required to encode the configuration data of the circuit-switched network).

V. System-Level Router Model

With increasing design complexities, there is a need for rapid design space exploration that makes use of a set of specifications. We model our NoC router framework using SystemC. By doing so, we functionally verify the model as well as setup a platform to estimate the advantages of this architecture over the baseline approach.

SystemC is a description language that abstracts the computation elements of a design by behaviours (or processes) and simplifies the communication between the cores using transaction level modelling. The framework has a set of library routines and macros implemented using C++. The behaviour of the hardware to be modelled is captured by simulating concurrent processes coded in C++.

SystemC Tool Flow: Every component in the router is modelled in C++ as a process. This .cpp file can be compiled and executed with the SystemC engine that is written in C++. We use the open source SystemC version 2.1 to compile our router design. The set of .cpp files are first compiled with the appropriate command options. Then, an executable is created to run the tool flow. We dump out the Value Change Dump (VCD) File from the engine.

The .VCD file of the router model can be used as follows:

• Applied to standard simulation tool for verifying the functionality of the model by viewing the waveform

• Estimate preliminary power consumed by the implementation on FPGAs, by using

Xpower and the architecture information (Virtex-4)

VI. Synthesis Results

In this section we present the Area/Synthesis results for our modified router implemented on Xilinx Virtex 4 [11]. The additional bandwidth offered by the proposed router comes with an increase in switch complexity. The amount of FPGA logic and routing resources consumed by the router instance depends on its complexity

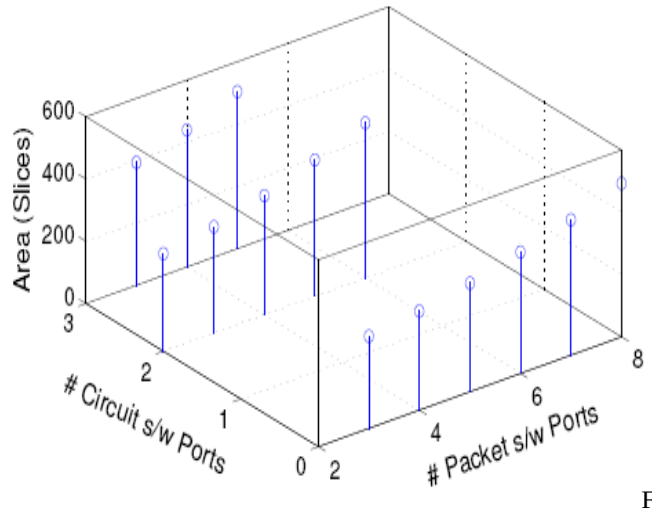


Figure 4: Design Parameters Vs Area

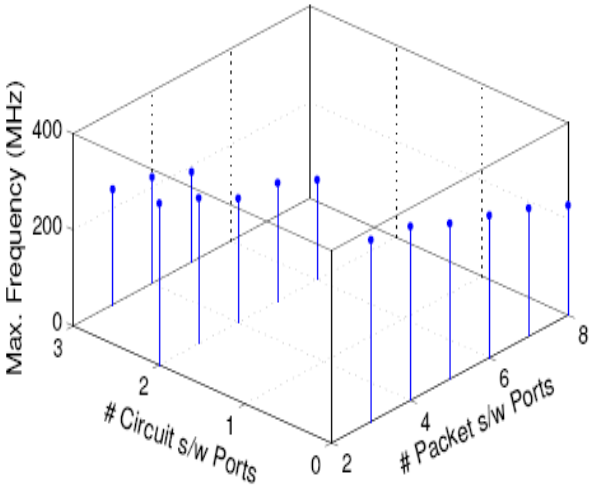


Figure 5: Design Parameters Vs Frequency

Figure 4 presents this variation in switch area with the number of ports (C & P-Layer) it supports. Further, the operating frequency of the router instances vary greatly due to different critical path lengths. Also, with increasing number of ports participating in the circuit-switched layer, the routing resources deplete rapidly (due to increased channel widths). This degradation in Performance in turn affects the bandwidth the switch can offer. Figure 5 presents the variation in switch operating frequency with the number of ports in both layers. The above area and frequency estimates are obtained by varying the parameters in the VHDL model of the router and by implementing them on the target device

Table 1: Scaling of Area and Frequency with No. of C-Layer Ports

MoClib Component	Area (Slices)	Frequency (MHz)
MC (4,2,2)	314	336
MC (5,3,2)	326	318
MC (5,2,3)	341	303
MC (6,3,3)	394	240
MC (6,2,4)	382	258
MC (7,3,4)	440	221

Furthermore, to perform automatic topology synthesis, we estimate the increase/decrease in switch area with exclusive variations in number of P-Layer ports and C-Layer ports independently. When NoC area is in the cost function, the above data will aid rapid design space exploration. Tables 1 and 2 present the scaling of area & frequency with increasing C-Layer and P-Layer ports respectively. In the tables, MC(x, y, z) denote an instance of the MoC lib library, where y is the total number of C-Layer ports, z is the total number of P-Layer ports and x is the sum of the two (total number of ports). Table 2 presents the scaling of area and frequency only with respect to the P-Layer ports and therefore they can be considered as variations of the MoCReS baseline router.

Table 2: Scaling of Area and Frequency with No. of P-Layer Ports

MoClib Component	Area (Slices)	Frequency (MHz)
MC (3,0,3)	296	378
MC (4,0,4)	318	362
MC (5,0,5)	349	324
MC (6,0,6)	390	296
MC (7,0,7)	435	267
MC (8,0,8)	493	229

VII. Results: Performance Improvement

Area vs Average Available Bandwidth/Port: The baseline version in this com- parson is MoCReS with 1VC+MC. The area (in slices) of the switch increases with the number of ports it supports. We measure the area values for increasing number of ports (packet-switched) in the baseline version. For similar area values, when the alternate hybrid router is used, there is an increase in available bandwidth per port. This band- width increase associated with the hybrid router architecture is compared in this section with the baseline approach. For equivalent area overheads (in slices) on a similar FPGA, Figure 6 presents the bandwidth capacity (in MB/s) of the NoC (per port) for both approaches. In spite of a rapid degradation in operating frequency (with increase in circuit-switched ports), there is a significant bandwidth gain using the hybrid two-layer approach. For the area window utilized in our library of routers, there is an average 20.4% gain in bandwidth (maximum of 24%) offered by our NoC. This gain in performance is due to supporting a high throughput circuit-switched layer with a marginal area overhead.

A. Design Issues

Even though it appears intuitively that an increase in number of ports in the C-layer gives performance benefits without any area overhead, there are certain design issues that can potentially limit the performance due to increase in switch complexity.

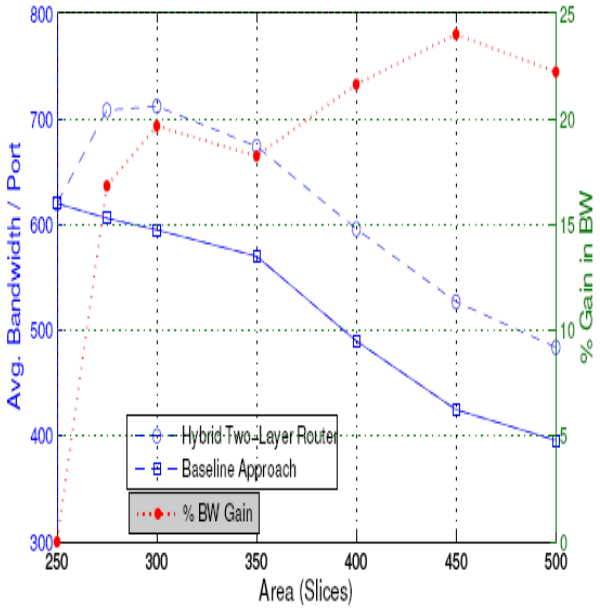


Figure 6: Area (Slices) Vs Avg. Bandwidth / Port

Operating frequency Vs Switch Complexity: There is a depletion of critical re- sources associated with an increase in switch complexity (number of ports, bus width). As a result, the operating frequency of the switch degrades which in turn affects the bandwidth offered by the router. For the NoC paradigm to efficiently be an alternative to the bus-based architecture, the performance design parameters must be chosen carefully so that it is possible to operate the routers at the highest possible frequency. Switch Power vs Link Power: By increasing the number of ports, we can reduce the average hop count [9], i.e we minimize the routers and links. This translates into a reduction in power consumed by the links, but an increase in power consumed by the switches. Beyond a cut-off, the increase in switch power can potentially overshadow the gain in link power; thereby it can increase the power/flit ratio

Explosion of Schedule Memory: With increasing number of C-layer ports, the schedule memory also scales linearly. The schedule memory, expressed in number of LUTs is a function of number of schedule cycles and C-layer ports present. If C is the number of ports participating in the C-Layer, then $\log_2 C$ is the number of configuration bits required per cycle.

Clock Signal Integrity: Operation of the C-layer ports requires the participating IP cores to be synchronous, as there is no buffering done, as opposed to packet-switch where multi-clock FIFOs separate the clock domains. Increasing the number of C-layer ports could potentially increase the distance between the connected IP cores. In this case, the signal integrity acts as a limitation to the number of C-layer ports, and reduces the clock rate.
It can be seen that all of the above factors limit the amount of performance gain that can be achieved using our hybrid approach. This trade-off between performance, area and port count merits a balance and requires an application-suitable tuning of the NoC topology.

VIII. Conclusions

To address the bandwidth limitations of MoCReS, we extend the design by developing hybrid two-layer router architecture. The novel design of the network component supports high throughput time-multiplexed circuit-switched connections between IPs interfaced to the same router, in addition to the packet-switched communication layer. Various instances of the NoC components are characterized for area and performance in the form of an MoC lib NoC component library. The advanced router architecture achieves an average improvement of 20.4% in NoC bandwidth (maximum of 24% compared to a traditional NoC).

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

REFERENCE

[1] Theodore Marescaux et al. Interconnectin Networks Enable Fine-Grain Multi-Tasking on FPGAs. In FPL'2002, pages 795–805, 2002.
[2] A.Kumar et al. An FPGA Design Flow for Reconfigurable Network-Based Multi-Processor Systems on Chip. In DATE'07, 2007.
[3] N.Kapre. Packet-Switched On-Chip FPGA Overlay Networks. M S thesis, California institute of technology. 2006.
[4] Clint Hilton and Brent Nelson. PNoC: a flexible circuit-switched NoC for FPGA based systems. In IEEE Proc. Computers and Digital Techniques, 2006.
[5] Manuel Saldaa, Lesley Shannon, and Paul Chow. The Routability of Multiprocessor Network Topologies in FPGAs. In SLIP'06, pages 49–56, 2006.
[6] T.A Bartic et. Al. Topology Adaptive Network-on-Chip Design and Implementation. In Computer and Digital Techniques, IEE Proceedings, pages 467–472, 2005.
[7] T.S.T. Mak et.al. On-FPGA Communication Architectures and Design Factors. In FPL'06, 2006.
[8] D. Wiklund and L.Dake. SoC BUS: switched network on chip for hard real time embedded systems. In Parallel and Distributed Processing Symposium, 2003.
[9] Balasubramanian Sethuraman and Ranga Vemuri. OptiMap: a tool for automated generation of NoC architectures using multi-port routers for FPGAs. In Design, Automation and Test in Europe, 2006. DATE '06, 2006.
[10] A.Janarthanan et.al. MoCReS: an Area-Efficient Multi Clock On-Chip Network for Reconfigurable Systems. In IEEE Computer Society ISVLSI'07, 2007.
[11] Xilinx Inc. http://www.xilinx.com.

AUTHORS PROFILE

Ezhumalai Periyathambi received the B.E degree in Computer Science and engineering from Madras University, Chennai , India in 1992 and Master Technology (M.Tech.,) in computer science and Engineering from J N T University, Hyderabad, India in 2006. He is currently working towards the Ph.D degree in Department of Information and Communication, Anna University, Chennai, India. He is working as assistant Professor in the Department of Computer Science and Engineering , Sri Venkateswara College of Engineering, sriperumbudur, Chennai, Tamilnadu, India. His research in reconfigurable architecture, Multi-Core Technology CAD – Algorithms for VLSI Architecture. Theoretical Computer Science. and mobile computing.

Arun Chokkalingam received the B.E degree in electronics and communication engineering from Bharathidasan University, Trichy , India in 2002 and the M.E degree from Anna University, Chennai, India 2004 and Doctorate in VLSI design at Anna University, Chennai TN, India in the year 2009. He is currently working towards the Ph.D degree in Department of Information and Communication, Anna University, Chennai, India. Since 2004 he has been an Lecturer in the Department of Information Technology, Sri Venkateswara College of Engineering, Chennai, Tamilnadu, India.His research in error correcting codes addresses effectively decoding algorithm and VLSI Architecture. His research interest including digital communication, coding theory, modulation and mobile communication.

S.Manoj kumar received the BE degree in Computer Science and Engineering from Bharadhidasan University, India in 2002 and Master of Engineering (ME) in Computer Science & Engineering from Anna University, Chennai, India in 2008. He is working as faculty in the Department of Computer Science and Engineering , Sri Venkateswara College of Engineering, sriperumbudur, Chennai, Tamilnadu, India. His research in reconfigurable architecture, Networking and mobile computing.

New System for Secure Cover File of Hidden Data in the Image Page within Executable File Using Statistical Steganography Techniques

Md. Rafiqul Islam, A.W. Naji, A.A.Zaidan and B.B.Zaidan*

Department of Electrical and Computer Engineering, Faculty of Engineering,
International Islamic University Malaysia (IIUM), P.O. Box 10, 50728 Kuala Lumpur, Malaysia

ABSTRACT— A **Previously traditional methods were sufficient to protect the information, since it is simplicity in the past does not need complicated methods but with the progress of information technology, it become easy to attack systems, and detection of encryption methods became necessary to find ways parallel with the differing methods used by hackers, so the embedding methods could be under surveillance from system managers in an organization that requires the high level of security. This fact requires researches on new hiding methods and cover objects which hidden information is embedded in. It is the result from the researches to embed information in executable files, but when will use the executable file for cover they have many challenges must be taken into consideration which is any changes made to the file will be firstly detected by untie viruses , secondly the functionality of the file is not still functioning. In this paper, a new information hiding system is presented. The aim of the proposed system is to hide information (data file) within image page of execution file (EXEfile) to make sure changes made to the file will not be detected by universe and the functionality of the exe.file is still functioning after hiding process. Meanwhile, since the cover file might be used to identify hiding information, the proposed system considers overcoming this dilemma by using the execution file as a cover file.**

(keyword): Information Hiding, portable executable file, Steganography, Statistical Technique.

I. INTRODUCTION

Information hiding is a general term encompassing many sub disciplines, is a term around a wide range of problems beyond that of embedding message in content. The term hiding here can refer to either making the information undetectable or keeping the existence of the information secret. Information hiding is a technique of hiding secret using redundant cover data such as images, audios, movies, documents, etc. This technique has recently become important in a number of application areas. For example, digital video, audio, and images are increasingly embedded with imperceptible marks, which may contain hidden signatures or watermarks that help to prevent unauthorized copy [1].It is a performance that inserts secret messages into a cover file, so that the existence of the messages is not apparent. Research in information hiding has tremendous increased during the past decade with commercial interests driving the field [1].

II. PORTABLE EXECUTABLE FILE (PE-FILE)

The Program Loader that is a subset of the Windows System assumes the loading executable files into a virtual memory, so the executable files have the format that the Program Loader can identify, and the format is called PE (Portable Executable). It is necessary to know the PE format and RVA which is an address type used in the PE in order to understand the new methods for hiding information in the PE, so we briefly describe the format[2],[3].

And the address type.The planned system uses a portable executable file as a cover to embed an executable program as an example for the planned system.This section is divided into four parts [4]:

- Characteristics of executable files.
- Techniques Related with PE-File.
- Executable files types.
- PE File Layout

A. Characteristics of Executable Files

The characteristics of the Executable file does not have a standard size, like other files, for example the image file (BMP) the size of this file is between (2-10 MB), Other example is the text file (TEXT) the size often is less than 2 MB.Through our study the characteristics of files have been used as a cover, it found that lacks sufficient size to serve as a cover for information to be hidden. For these features of the Executable file, it has unspecified size; it can be 650 MB like window setup File or 12 MB such as installation file of multi-media players. For taking advantage of this feature (disparity size) make it a suitable environment for concealing information without detect the file from attacker and discover hidden information in this file[3].

B. Techniques Related with PE-File.

- RVA

RVA is a position unit in the PE, and the RVA is used as an offset from the start-address of a PE file loaded on the memory. The start-address of a file on the memory is in Image Base that is one of the attributes of the PE file. For instance, if the Image Base of a file is 0x00400000 and one position of the file is 0x1000(RVA), the position on the memory will be 0x00401000[3],[4].

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

- PE Format

The header of PE format starts with MS-DOS stub that is used for printing a message, "This program cannot be run in DOS mode", if the operating system can't identify the PE on execution time.IMAGE_NT_HEADER located in the position after the MS-DOS stub has the information for the execution of a file, and consists of IMAGE_FILE_HEADER and IMAGE_OPTIONAL_HEADER. The IMAGE_FILE_HEADER has the information on the file, such as create time and machine type. The IMAGE_OPTIONAL_HEADER has the information on functions used in the file and on the start-address of the file on a memory, and the infor-mation is managed by IMAGE_DATA_DIRECTORY. A PE file except the header is composed of several sections that are basic unit of code or data within a PE or COFF file. IMAGE_SECTION_HEADER that is located in the position following to IMAGE_ OPTIONAL_HEADER has the information on each section. The information consists of PointerToRawData, SizeOfRawData, VirtualAddress, and VirtualSize. The PointerToRawData and the SizeOfRawData respectively mean the position of each section and the size of each section on the file. The VirtualAddress and the VirtualSize respectively mean the position of each section and the size of each section on the memory. The size of each section on the file is a multiple of FileAlignment that is in IMAGE_OPTIONAL_HEADER. If the amount of the data of a section is smaller than the size of the section that is allotted on compile time, the slack space of the section occurs. The common sections used in the PE include a .text that has program binaries, .data, .idata that has information on export and import functions, .edata, and .rsrc section. An .idata section has the information on import functions used in executable files during the period of an execution [5],[6].

C. *Executable File Types*

The number of different executable file types is as many and varied as the number of different image and sound file formats. Every operating system seems to have several executable file types unique to it. These types are [4],[5],[6]:

- EXE (DOS"MZ")

DOS-MZ was introduced with MS-DOS (not DOS v1 though) as a companion to the simplified DOS COM file format. DOS-MZ was designed to be run in real mode and having a relocation table of SEGMENT: OFFSET pairing. A very simple format that can be run at any offset, it does not distinguish between TEXT, DATA and BSS.The maximum file size of (code + data + bss) is one-mega bytes in size. Operating systems that use are: DOS, Win*, Linux DOS.

- EXE (win 3.xx "NE"):

The WIN-NE executable formatted designed for windows 3.x is the "NE" new-executable. Again, a 16-bit format, it alleviates the maximum size restrictions that the DOZ-MZ has. Operating system that uses it is: windows 3.xx.

- EXE (OS/2 "LE"):

The "LE" linear executable format was designed for IBM's OS/2 operating system by Microsoft supporting both 16 and 32-bit segments operating systems that are used in: OS/2, DOS.

- EXE (win 9x/NT "PE"):

With windows 95/NT a new executable file type is required, thus was born the "PE" portable executable. Unlike its predecessors, the WIN-PE is a true 32-bit file format, supporting releasable code. It does distinguish between TEXT, DATA, and BSS. It is in fact, a bastardized version of the common object file format (COFF) format. Operating systems that use it are: windows 95/98/NT/2000/ME/CE/XP.

- ELF:

The ELF, Executable Linkable Format was designed by SUN for use in their UNIX clone. A very versatile file format, it was later picked up by many other operating systems for use as both executable files and as shared library files. It does distinguish between TEXT, DATA and BSS.
TEXT: the actual executable code area.
DATA: "initialized" data, (Global Variables).
BSS : "un- initialized" data, (Local Variables).

D. *PE File Layout*

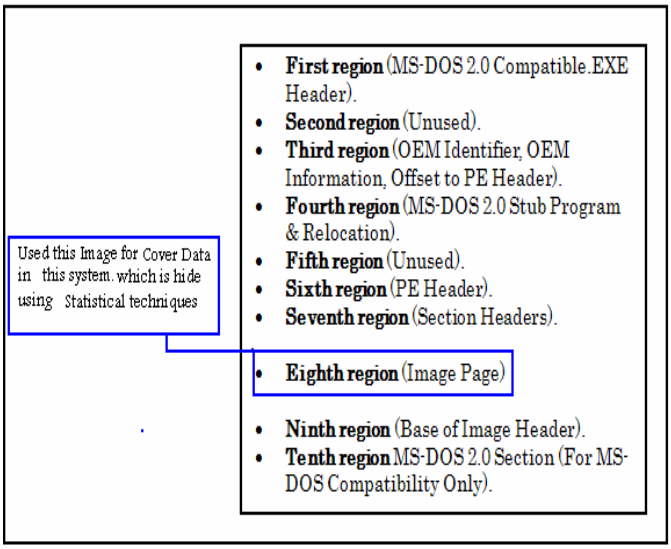


Figure 1.Typical 32-bit Portable .EXE File Layout.

III. STEGANOGRAPHY

Steganography is the art and science of writing hidden messages in such a way that no-one can realizes there is a hidden message in data (e.g. images file, documents file, sounds file … etc) except the sender and intended recipient.The word steganography is of Greek origin and means "covered, or hidden writing" .Cryptography obscures the meaning of a message, but it

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

does not conceal the fact that there is a message, cryptography is the art of secret writing, which is intended to make a message unreadable by a third party but does not hide the message in a communication [1],[5]. Although steganography is separate and distinct from cryptography, but there are many analogies between the two, and some authors categorize steganography as a form of cryptography since hidden communication is a form of secret writing [2],[6]. A Watermark is a recognizable image or pattern in paper that appears as various shades of lightness/darkness when viewed by transmitted light (or when viewed by reflected light, atop a dark background), caused by thickness variations in the paper ,Digital Watermarking is the process of embedding information into a digital signal. The signal may be audio, pictures or video, for example. If the signal is copied, then the information is also carried in the copy. Steganography and digital watermarking are the same but the purpose of last one is use to copyright protection systems, which are intended to prevent or deter unauthorized copying of digital media. Steganography is an application of digital watermarking, where two parties communicate a secret message embedded in the digital signal [2].Classical Steganography often used methods of completely obscuring the message so it was unnoticeable to those who didn't know the specific covert method it was using for example Invisible Inks, which was able to write a confidential letter with any other non-value-confidential and usually write between lines. Modern Steganography refers to hide information in digital picture, audio or text files …etc , each one of this digitals data has a many techniques can use with it , for example in digital image the JPHide/JPSeek uses the coefficients in a JPEG to hide information, this method alter the image. In digital audio file several packages also exist for hiding data in audio files, Such as MP3Stego not only effectively hides arbitrary information, but also claims to be a partly robust method of watermarking MP3 audio files [4]. The Windows Wave format lets users hide data using Steghide, it alters the least significant bits (LSB) of data in the carrier medium [3].In a growing number of applications like digital rights management, covert communications, hiding executables for access control, annotation etc .all these application scenarios given the multimedia steganography techniques have to satisfy two basic requirements. The first requirement is perceptual transparency or noticeable perceptual distortion, i.e. cover object (object not containing any additional data) and stego object (object containing secret message) must be perceptually indiscernible [3]. The second constraint is high data rate of the embedded data. All the stego-applications, besides requiring a high bit rate of the embedded data, have need of algorithms that detect and decode hidden bits without access to the original multimedia sequence (blind detection algorithm) [4].

A. Characterization of Steganography Systems

Steganographic techniques embed a message inside a cover. Various features characterize the strength and weaknesses of the methods. The relative importance of each feature depends on the application [7].

- Capacity

The notion of capacity in data hiding indicates the total number of bits hidden and successfully recovered by the Stego system.

- Robustness

Robustness refers to the ability of the embedded data to remain intact if the stego-system undergoes transformation, such as linear and non-linear filtering; addition of random noise; and scaling, rotation, and loose compression.

- Undetectable

The embedded algorithm is undetectable if the image with the embedded message is consistent with a model of the source from which images are drawn. For example, if a Steganography method uses the noise component of digital images to embed a secret message, it should do so while not making statistical changes to the noise in the carrier.Undetectability is directly affected by the size of the secret message and the format of the content of the cover image.

- Invisibility (Perceptual Transparency)

This concept is based on the properties of the human visual system or the human audio system. The embedded information is imperceptible if an average human subject is unable to distinguish between carriers that do contain hidden information and those that do not. (Ross, 2005) It is important that the embedding occurs without a significant degradation or loss of perceptual quality of the cover.

- Security

It is said that the embedded algorithm is secure if the embedded information is not subject to removal after being discovered by the attacker and it depends on the total information about the embedded algorithm and secret key.

B. Statistical Steganography Techniques

Statistical steganography techniques utilize the existence of "1-bits" Steganography schemes, which embed one bit of information in a digital carrier. This is done by modifying the cover in such a way that some statistical characteristics change significantly if a "1" is transmitted. Otherwise, the cover is left UN changed. So the receiver must be able to distinguish unmodified covers from modified ones. A cover is divided into l (m) disjoint blocks B1...B l (m)[8]. A secret bit, mi is inserted into the ith block by placing "1" in to Bi if mi=1.Otherwise, the block is not changed in the embedding process.The detection of a specific bit is done via a test function which distinguishes modified block from unmodified block (1)[1],[8]:

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

$$f(B_i) = \begin{cases} 1 & \text{block } B_i \text{ was modified in the embedding process} \\ 0 & \text{otherwise} \end{cases}$$

.. .. (1)

The function f can be interpreted as a hypothesis-test function and the test of null-hypothesis "block Bi was not modified "against the alternative hypothesis "block Bi was modified." Therefore, the whole class of such steganography systems statistical steganography .the receiver successively applies f to alaa cover-block Bi in order to restore every bit of the secret message. The main question which remains to be solved is how such a function f in (8) can be constructed. If they interpret f as a hypothesis-testing function , they can use the theory of hypothesis testing from mathematical statistics .Let us assume could find a formula h(Bi), which depends on some elements of the cover-block Bi, and knew the distribution of h(Bi), in the unmodified block (i.e,the Hypothesis holds in this case) could then use standard procedure to test if h(Bi), equals or exceeds a specific value. If managed to alter h(Bi) in the embedding process in a way that its expected value is 0 if the block Bi was not modified, and expected value is much greater otherwise , could test whether h(Bi) equals zero under the given distribution of h(Bi). Statistical steganography techniques are, however, difficult to apply in many cases. First, a good test statistic h(Bi) must be found which allows distinction between modified and unmodified cover-blocks. Additionally, the distribution of h (Bi) must be known for a "normal" cover; in most cases, this is quite a difficult task. In practical implementations many (quite questionable) assumptions are made in order to determine a closed formula for this distribution. As an example, wanted to construct a statistical steganography algorithem out of pitas' watermarking system, which is similar the patchwork approach of bender et al. Suppose every cover-block Bi is a rectangular set of pixels p(i)n,m .Furthermore, let S={s(i)n,m} be a rectangular pseudorandom binary pattern of equal size, where the number of one is S equals the number of zeros. Would assume that both the sender and receiver have access to S, which represents the stego-key in this application. The sender first splits the image block Bi into two sets, Ci and Di of equal size (i.e., putting all pixels with indices (n,m)into set C where the corresponding key bit n,m equals zero)[1],[8]:

$$\begin{aligned} C_i &= \{p^{(i)}_{n,m} \in B_i | s_{n,m} = 1\} \\ D_i &= \{p^{(i)}_{n,m} \in B_i | s_{n,m} = 0\} \end{aligned}$$

… … (2)

The sender then adds a value k > 0 to all pixels in the subset Ci but leaves all pixels in Di unchanged. In the last step, Ci and Di are merged to form the marked image block Bi. In order to extract the mark, the receiver reconstructs the sets Ci and Di. If the block contains a mark, all value in Ci will be larger than the corresponding values in the embedding step; thus testing the difference of the means of sets Ci and Di. If assumed that all pixels in both Ci and Di are independent identically distributed random variables with an arbitrary distribution, the test statistic [1],[8]:

$$q_i = \frac{\overline{C_i} - \overline{D_i}}{\hat{\sigma}_i}$$

with

$$\hat{\sigma}_i = \sqrt{\frac{\mathrm{Var}[C_i] + \mathrm{Var}[D_i]}{|S|/2}}$$

… ... (3)

Where $\overline{Ci}$ denotes the mean over all pixels in the set Ci and Var [Ci] the estimated variance of the random variables in Ci , will follow a N(0,1) normal distribution asymptotically due to the central limit theorem . if a mark is embedded in the image block Bi , the expected value of q will be greater than zero. The receiver is thus able to reconstruct the ith secret message bit by testing whether the statistic qi of block Bi equals zero under the N (0, 1) distribution[7],[8].

IV. METHODOLOGY

A. System Overview

The most important reason behind the idea of this system is that the programmers always need to create a back door for all of their developed applications, as a solution to many problems such that forgetting the password. This idea leads the customers to feel that all programmers have the ability to hack their system any time. At the end of this discussion all customers always are used to employ trusted programmers to build their own application. Programmers want their application to be safe anywhere without the need to build ethic relations with their customers. In this system a solution is suggested for this problem. The solution is to hide the password in the executable file of the same system and then other application to be retracted by the customer himself. Steganography needs to know all files format to find a way for hiding information in those files. This technique is difficult because there are always large numbers of the file format and some of them have no way to hide information in them.

B. System Concept

Concept of this system can be summarized as hiding the password or any information beyond the end of an executable file so there is no function or routine (open-file, read, write, and close-file) in the operating system to extract it. This operation can be performed in two alternative methods:

- Building the file handling procedure independently of the operating system file handling routines. In this case we need canceling the existing file handling routines and developing a new function which can perform our need, with the same names. This way needs the customer to install the system application manually as shown in Figure 2.

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

- Developing the file handling functions depending on the existing file handling routines. This way can be performed remotely as shown in Figure 3.

The advantage of the first method is it doesn't need any additional functions, which can be identified by the analysts. The disadvantage of this method is it needs to be installed (can not be operated remotely). The advantage of the second method is it can be executed remotely and suitable for networks and the internet applications. So we choose this concept to implementation in this paper.

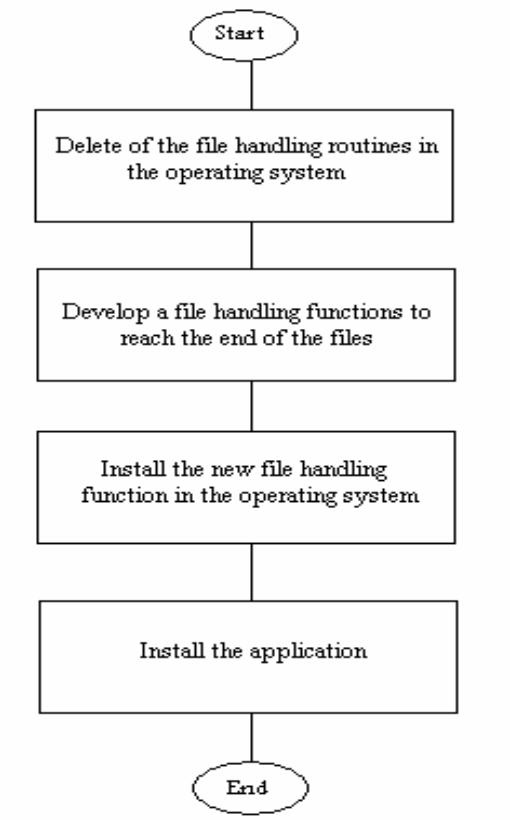


Figure 2. First Method of the System Concept

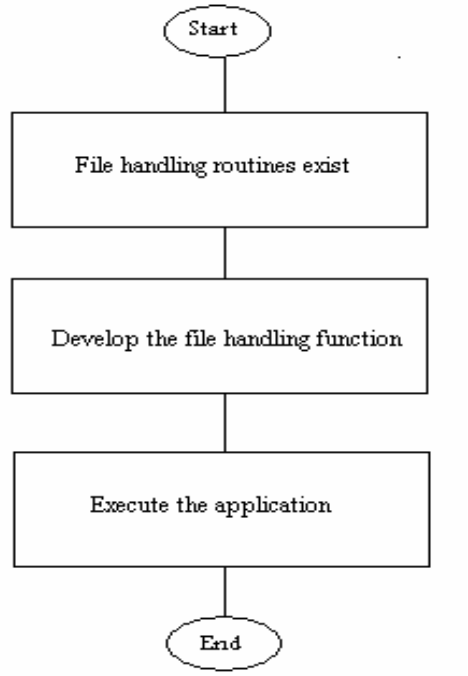


Figure 3. Second Method of the System Concept

C. System Features

This system has the following feature*:*

- The hiding operation within image page of EXE file using the statistical technique increases the degree of security for the information hiding which is used in the proposed system because the data which is embedded inside the EXE file is not embed directly of EXE file , it will be hiding within image page of EXE file. So the attacker can not be guessing the information hidden.

- The cover file can be executed normally after hiding operation. Because the hidden information already hide in the image page within exe.file and thus cannot be manipulated as the exe.file, therefore, the cover file still natural, working normally and not effected, such as if the cover is EXE file (WINDOWES XP SETUP) after hiding operation it'll continued working, In other words, the EXE file can be installed of windows.

- Virus detection programmers' can't detect such as files, the principle of antivirus check are checking from beginning to end. When checking the exe.files by antivirus, will checked it from beginning to end of it, since the principle of information hiding for this system within image page of EXE file .The information hiding will be hide inside the image page and the EXE file after hiding process is same manufacture of EXE file before hiding process. That is why the EXE file undetectable by Unit-Virus.

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

D. The Proposed System Structure

The system has been implementation by using Java. The block flow of hiding operation can be performed as shown in Figure 4. The following algorithm is the hiding operation procedure. The block flow of retract operation can be performed as shown in Figure 5. The following algorithm is the retract operation procedure.

The following algorithm is the hiding operation procedure:

Procedure: Hide operation.
Input: Hidden file name, cover file name.
Output: Stego-File.

- Begin (1).
- Opens the cover file (EXE file).
- Assign a pointer to the end of (Section header), which is before image page of the cover file.
- Select the image page consider normal page, it consider the cover for data.
  - Begin (2) for the image page.
  - Write the hidden file name.
  - Assign a pointer after hidden file name.
  - Write the hidden file content.
  - End(2) for the image page
- End (1).

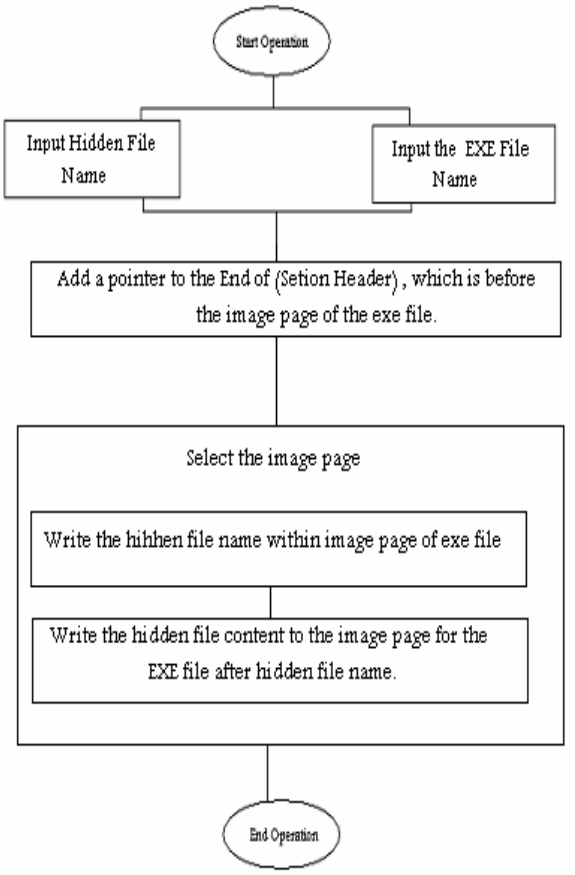


Figure 4. Block Flow of Hiding Operation.

The following algorithm is the retract operation procedure:

- Begin (1).
- Select the cover file.
- Get the End of the (Section Header) of EXE File.
- Select the image page:
  - Begin (2) for the image page.
  - Read the name of the hidden file.
  - Read the Hidden data.
  - Create a file using hiding file name.
  - Write in to the Create file the hiding data.
  - End (2) for the image page.
- End(1)

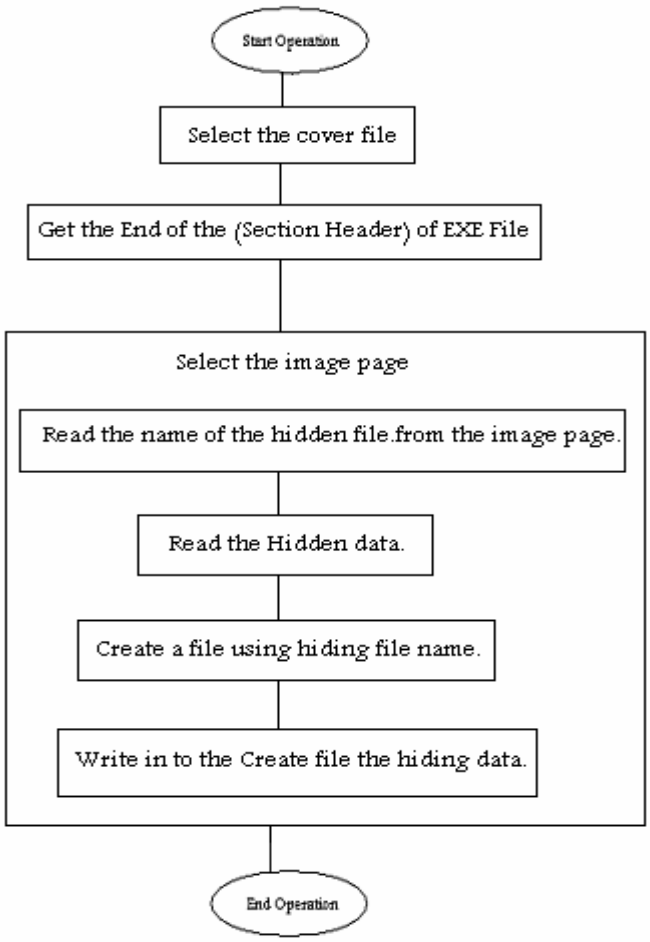


Figure 5. Block Flow of Retract Operation

V. CONCLUSION

The .EXE files are one of the most important files in operating systems and in most systems designed by developers (programmers/software engineers), and then hiding information in these file is the basic goal for this paper, because most users of any system cannot alter or modify the content of these files. We get the following conclusions:

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

- PE files structure is very complex because they depend on multi headers and addressing, and then insertion of data to PE files without full understanding of their structure may damage them, so the choice is to hide the information beyond the structure of these files, so the approach of the proposed system is to prevent the hidden information to observation of these systems.
- One of important conclusion most antivirus systems do not allow direct write in executable file, so the approach of the proposed system is to prevent the hidden information to observation of these systems.
- The cover file can be executed normally after hiding operation. Other word the cover file still natural, working normally and not affected.

VI. FUTURE WORK

There are many suggestions for improving the proposed system, the main suggestions are:

- Developing the method which is used in proposed system to deal with other PE files such as "dll", "sys", "cpl", and "ocx".
- Developing the proposed system to deal with other executable files created by other operating systems like (LINUX, UNIX or OS/2).

REFERENCES

[1] A.A.Zaidan, B.B.Zaidan, Fazidah Othman, "New Technique of Hidden Data in PE-File with in Unused Area One", International Journal of Computer and Electrical Engineering (IJCEE), Vol.1, No.5, ISSN: 1793-8198, pp 669-678.

[2] A.A.Zaidan, B.B.Zaidan, M.M.Abdulrazzaq, R.Z.Raji, and S.M.Mohammed," Implementation Stage for High Securing Cover-File of Hidden Data Using Computation Between Cryptography and Steganography", International Conference on Computer Engineering and Applications (ICCEA09), Telecom Technology and Applications (TTA), Vol.19, Session 6, p.p 482-489, ISBN: 978-1-84626-017-9,June 6 (2009), Manila, Philippines

[3] Alaa Taqa, A.A Zaidan, B.B Zaidan ,"New Framework for High Secure Data Hidden in the MPEG Using AES Encryption Algorithm", International Journal of Computer and Electrical Engineering (IJCEE),Vol.1 ,No.5, ISSN: 1793-8198, pp.589-595 .

[4] A.W.Naji, A.A.Zaidan, B.B.Zaidan, Shihab A, Othman O. Khalifa, " Novel Approach of Hidden Data in the (Unused Area 2 within EXE File) Using Computation Between Cryptography and Steganography ", International Journal of Computer Science and Network Security (IJCSNS) , Vol.9, No.5 , ISSN : 1738-7906, pp. 294-300.

[5] A.W. Naji, A.A.Zaidan, B.B.Zaidan, Ibrahim A.S.Muhamadi, "Novel Approach for Cover File of Hidden Data in the Unused Area Two within EXE File Using Distortion Techniques and Advance Encryption Standard.", Academic and Scientific Research Organizations (WASET), International Conference on Computer, Electrical, and Systems Science, and Engineering (CCESSE09), , ISSN:2070-3724, 26-28 .

[6] A.W. Naji, A.A.Zaidan, B.B.Zaidan, Ibrahim A.S.Muhamadi, "New Approach of Hidden Data in the portable Executable File without Change the Size of Carrier File Using Distortion Techniques", Academic and Scientific Research Organizations (WASET), International Conference on Computer, Electrical, and Systems Science, and Engineering(CCESSE09), , ISSN:2070-3724.

[7] A.W. Naji,Teddy S. Gunawan and Shihab A. Hameed, B.B Zaidan, A.A Zaidan " "Stego-Analysis Chain, Session One" Investigations on Steganography Weakness Vs Stego-Analysis System for Multimedia File ", International Conference on IACSIT Spring Conference (IACSIT-SC09), Advanced Management Science (AMS), Listed in IEEE Xplore Session 9, P.P 393-397 , ISBN:978-7695-3653-8, April 17 (2009) , Singapore.

[8] B.B.Zaidan, A.A.Zaidan, Fazidah. Othman, Ali Rahem," Novel Approach of Hidden Data in the (Unused Area 1 within EXE File) Using Computation Between Cryptography and Steganography ", Academic and Scientific Research Organizations (WASET), International Conference on Cryptography, Coding and Information Security (ICCCIS09), Vol.41, Session 24, ISSN: 2070-3740.

Author Information

Md. Rafiqul Islam received his B Sc (Electrical and Electronic Engineering) from BUET, Dhaka in 1987. He received his MSc and PhD both in Electrical Engineering from UTM in 1996 and 2000 respectively. He is Fellow of IEB and member of IEEE. He is currently faculty member of Electrical and Computer Engineering Department of International Islamic University Malaysia. His area of research interest are radio link design, RF propagation measurement and RF design, smart antennas and array antennas design etc.

Dr. Ahmed Wathik Naji - He obtained his 1st Class Master degree in Computer Engineering from University Putra Malaysia followed by PhD in Communication Engineering also from University Putra Malaysia. He supervised many postgraduate students and led many funded research projects with more than 50 international papers. He has more than 10 years of industrial and educational experience. He is currently Senior Assistant Professor, Department of Electrical and Computer Engineering, International Islamic University Malaya, Kuala Lumpur, Malaysia.

Aos Alaa Zaidan He obtained his 1st Class Bachelor degree in Computer Engineering from university of Technology / Baghdad followed by master in data communication and computer network from University of Malaya. He led or member for many funded research projects and He has published more than 40 papers at various international and national conferences and journals, currently he is working on the multi module for Steganography, Development & Implement a novel Skin Detector. He is members IAENG, WASET, and IACSIT.

Bilal Bahaa Zaidan He obtained his bachelor degree in Mathematics and Computer Application from Saddam University/Baghdad followed by master from Department of Computer System & Technology Department Faculty of Computer Science and Information Technology/University of Malaya /Kuala Lumpur/Malaysia, He led or member for many funded research projects and He has published more than 40 papers at various international and national conferences and journals. He is members IAENG, WASET, and IACSIT.

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

AN INNOVATIVE PLATFORM TO IMPROVE THE PERFORMANCE OF EXACT-STRING-MATCHING ALGORITHMS

Mosleh M. Abu-Alhaj[1], M. Halaiyqah[2], Muhannad A. Abu-Hashem[2], Adnan A. Hnaif[1], O. Abouabdalla[1,] and Ahmed M. Manasrah.
[1]: National Advanced IPv6 Center of Excellence, [2]: Computer Science
University Sains Malaysia, Penang Malaysia

ABSTRACT

Exact-String-Matching is an essential issue in many computer science applications. Unfortunately, the performance of Exact-String-Matching algorithms, namely, executing time, does not address the needs of these applications. This paper proposes a general platform for improving the existing Exact-String-Matching algorithms executing time, called the PXSMAlg platform. The function of this platform is to parallelize the Exact-String-Matching algorithms using the MPI model over the Master/Slaves paradigms. The PXSMAlg platform parallelization process is done by dividing the Text into several parts and working on these parts simultaneously. This improves the executing time of the Exact-String-Matching algorithms. We have simulated the PXSMAlg platform in order to show its competence, through applying the Quick Search algorithm on the PXSMAlg platform. The simulation result showed significant improvement in the Quick Search executing time, and therefore extreme competence in the PXSMAlg platform.

Keywords- String matching, Parallel, Quick search

I. INTRODUCTION

Computer science applications play a significant role in many fields, such as DNA analysis, artificial intelligence, and information retrieval, among various others. String matching is an important issue in many of these applications. It is the process of finding the occurrence of a Pattern P into a Text T, wherein T is longer than P. This occurrence is either exactly matched or partially matched with the Pattern. Accordingly, string matching algorithms are divided into two main categories: Exact-String-Matching algorithms and approximate string matching algorithms. Exact-string-matching algorithms are concerned with the number of occurrences of the pattern into a given text, while approximate string matching algorithms are concerned with the similarity percentage between the pattern and the text or any part of the text [1] [2]. This paper concentrates on Exact-String-Matching algorithms, such as the Boyer-Moore, Horspool, and Quick Search algorithms [3].

Currently, the world is witnessing a revolution in hardware efficiency, where a normal laptop can have a multi-core processor. To take advantage of this revolution, most of the applications are used in parallel computing, wherein a problem is divided into smaller problems, which are then processed simultaneously. Moreover, many parallel paradigms and models have been developed and proposed. The Master/Slave paradigm is a widely used paradigm in parallel computing. It is a Multi-Processors paradigm containing several nodes, one node is the master and the other nodes are the slaves. The master node is responsible for maintaining global data structures and partitioning the overall computational problem into smaller sub–problems, which are handed to the slaves to process for computation. On the other hand, the Message Passing Interface (MPI) is one of the well-known parallel models used in parallel computing above the hardware and memory architectures. In this paper, we will use the MPI model along with the Master/Slave paradigm to develop a general parallel platform and improve the Exact-String-Matching algorithms' performance [4] [5] [6].

I.1. Quick search algorithm

Sunday [7] proposed and designed a new algorithm for string matching, which is faster than the Boyer-moor algorithm and is considered one of the fastest algorithms in the string matching field. Its time and space complexity are O(m + n) and O (n), respectively. In terms of detecting matches between two strings, the quick search algorithm looks similar to the Boyer-moor algorithm. However, the difference between them is that the quick search algorithm only uses the bad-character shift table while the Boyer-Moore uses both bad-character shift and good suffix shift tables. Moreover, this algorithm starts searching from the left-most character to the right [7].

The rest of this paper is arranged as follows. Section 2 discusses some of the related works. Section 3 discusses the proposed platform, highlights the border problem, and shows the proposed platform performance. Finally, the conclusion is stated in Section 4.

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

II. RELATED WORKS

There have been several research works on parallel Extract-String-Matching algorithms. For example, Raju and Babu [8] proposed a parallel technique for string matching algorithm. They considered the linear array with a reconfigurable pipelined bus system (LARPBS) and 2D LARPBS for string matching in their work, which has many existing applications such as cellular automata, computational biology, and string database. The proposed method introduced increases the speedup of the string matching process using LARPBS. They obtained time complexity O (1) for the string matching on 2D LARPBS where no preprocessing is done to the text and the pattern [8].

Park and George [9] presented a dataflow schemes string matching algorithms parallelization. In their work, they covered exact matching and k-mismatched problems, which they consider as sub-problems in the string matching field. The time complexity of the proposed parallel algorithm was $O((n/d)+\alpha)$, $0 \le \alpha \le m$, where n and m are the length of the text and pattern with (n >> m) and d is the number of streams used. The parallelism degree can be controlled by changing the value of the variable d, which is present in the input streams. Due to the one-pass dataflow algorithms, there was no preprocessing and memory space used for this schema [9].

III. PARALLEL-EXACT-STRINGS-MATCHING-ALGORITHM

Exact-String-Matching is one of the main problems in many computer applications. One of the Exact-String-Matching problems is the slow matching process between the Pattern and the Text. Parallel computing is a key technique used to reduce the time of the Exact-String-Matching process. In this paper, we have exploited one of the Parallel computing models, namely, the MPI model, in order to provide a general platform to parallelize the Exact-String-Matching algorithms. The proposed platform, called Parallel-Exact-Strings-Matching algorithm (PXSMAlg), can be applied in all the Exact-String-Matching algorithms, such as Quick Search. The PXSMAlg platform has been developed to run the Master/Slave paradigm. [3] [5] [6].

III.1. The PXSMAlg Platform Process

The parallelization process of the PXSMAlg platform is accomplished through a set of steps. First, the Master node reads the Pattern and the Text (Source-File). Second, the Master node calculates the Source-File size and divides it into multiple parts, according to the determined nodes number. Then, the Master node distributes each part to a specific node. After that, the searching function starts in each node to find the Pattern, with each node searching in its source file part. Before the final step is done, each node checks the border of its neighbor, except for the last node. The border issue will discussed later. Finally, the number of matches is collected from all nodes. Meanwhile, the Master node calculates all the collected results and then prints the total result. Figure 1 illustrates the PXSMAlg platform.

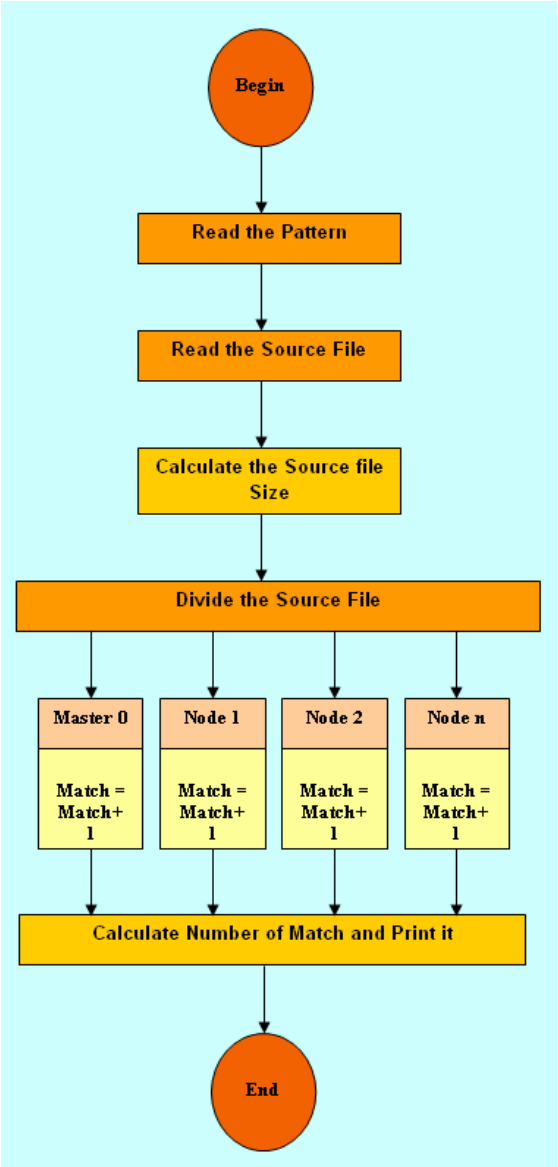


Figure 1: The PXSMAlg platform process

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

III.2. Handling the Nodes Borders Issue

Border issue is key element that the PXSMAlg platform faces. This issue happened when the Pattern located between the Source-File parts borders led to a mismatched (i.e., not found) Pattern. To resolve this issue, the PXSMAlg platform allows node "n" to check the border between node "n" and node "n+1," node "n+1" to check the border between node "n+1" and node "n+2," and so on. For illustrative purposes, suppose the Source-File is "EXACT STRINGS MATCHING," the Pattern is "INGS," and the number of nodes is two. First, divide the Source-File into two parts according to the number of nodes, Part1 is "EXACT STRIN" and Part2 is "GS MATCHING." As we can notice, if node1 searched for the Pattern "INGS" in the border of the two parts, it will find it; otherwise, node1 will not be able to find the Pattern "INGS" in Part1 and node2 will not be able to find the Pattern "INGS" in Part2.

III.3. PXSMAlg Platform Performance Analysis

We have built a simulation to demonstrate the feasibility of the PXSMAlg platform and its compatibility with the Exact-String-Matching algorithms. In addition, this simulation is done to compare the performance of the PXSMAlg platform with the conventional method, that is, the sequential method. The simulation built is based on three main factors: executing time, speedup, and efficiency. Our simulation runs under the Aurora server, which consists of 14 nodes, with each node having 2 CPUs, a speed of 1300MHz and a 1GB memory; all nodes run the Linux OS. The results showed high performance of the PXSMAlg platform over the sequential methods.

We have carried out 14 different experiments to search for the letter “a” in a 37 MB file size. We have applied the experiments using the Quick Search algorithm, which is one of the best algorithms in the Exact-Strings-Matching algorithms. The result showed significant improvement in the executing time and speedup, wherein applying the Exact-String-Matching algorithms on the PXSMAlg platform decreased the executing time, especially when compared with the sequential executing time. Figure 2 depicts the improvement in the Quick Search algorithm process time in the sequential mode, one node, parallel mode, and two or more nodes. In addition, the speedup is increased by applying the Exact-String-Matching algorithms on the PXSMAlg platform. Figure 3 shows the improvement in speedup in the Quick Search algorithm. In contrast to the executing time and the speedup, the processors’ efficiency decreases by applying Exact-String-Matching algorithms on the PXSMAlg platform. Figure 4 shows the decreasing efficiency in the Quick Search algorithm when the number of the processors increases.

IV. CONCLUSION

In this paper, we have proposed a general platform, called the PXSMAlg platform, in order to improve the Exact-Strings-Matching algorithms performance. The PXSMAlg platform relies on using the MPI model over the Master/Slave Paradigm to improve the Exact-Strings-Matching algorithms’ competence in terms of speeding up the executing time. We have applied one of the best Exact-String-Matching algorithms, the Quick Search algorithm, on the PXSMAlg platform. The result showed high efficiency in the PXSMAlg platform. In comparison with the sequential mode, the Quick Search executing time and speedup were highly improved. On the other hand, the efficiency of the processors decreased.

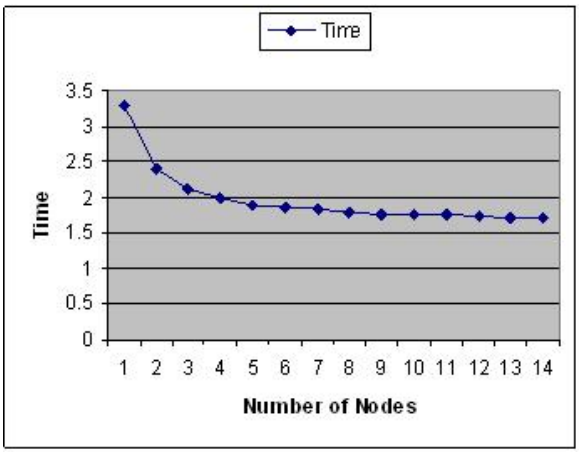


Figure 2: Executing Time

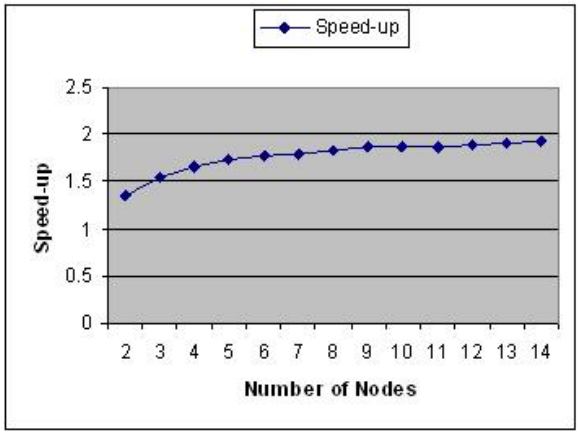


Figure 3: Speedup

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

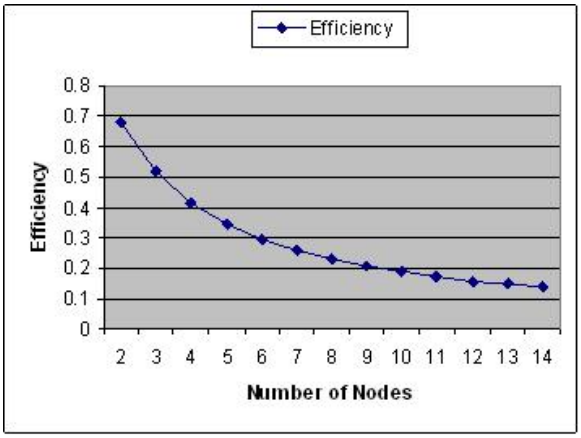


Figure 4: Efficiency

REFERENCES

[1] BERMAN, K. A. & PAUL, J. L., Algorithms: Sequential, Parallel and Distributed, Thomson, United State of America., 2005.

[2] S.Viswanadha Raju and A.Vinaya Babu, ”Optimal Parallel algorithm for String Matching on Mesh Network Structure”, International Journal applied mathematica Sciences, Vol. 3 No.2, 167-175, 2006.

[3] ALTSCHUL, S. F., MADDEN, T. L., SCHAFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. J., "Gapped PLAST and PSI-BLAST : A New Generation of Protein Database Search Programs. Nucleic Acids Research", Vol. 25, No. 17, pp 3389-3402, 1997.

[4] http://web.it.kth.se/~matsbror/multicore/, Feb 2009.

[5] C.H. Hsu, T.L. Chen, and G.H. Lin, “Grid Enabled Master Slave Task Scheduling for Heterogeneous Processor Paradigm,” Proc. Fourth Int’l Conf. Grid and Cooperative Computing (GCC ’05), pp. 449-454, 2005.

[6] Blaise Barne and, Lawrence Livermore, "https://computing.llnl.gov/tutorials/parallel_comp/", Jan 2009.

[7] D.M. Sunday, "A very fast substring search algorithm", Comm. ACM, No. 33, pp 132–142, pp 1990.

[8] S. Viswanadha Raju, & A. Vinaya Babu, "Parallel algorithms for string matching problem on single and two dimensional reconfigurable pipelined bus systems", Journal of Computer Science Vol. 3, No.9, pp.754-759, September 2007.

[9] PARK, J. H., AND GEORGE, K. M., "Parallel string matching algorithms based on Dataflow", In Proceedings of the 32rd Hawaii International Conference on System Sciences, 1999.

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

A MAC Layer Based Defense Architecture for Reduction-of-Quality (RoQ) Attacks in Wireless LAN

Jatinder Singh,
Director, Universal Institute of
Engg. & Tech. Lalru-CHD(India)

Dr. Savita Gupta,
Prof. Deptt. Of Computer Engg.
UIET, PunjabUniversity, CHD (India (India)

Dr. Lakhwinder Kaur,
Reader, UCOE, Punjabi) University,Patiala

Abstract

Recently an alternative of DDoS attacks called shrew attacks or Reduction-of-Quality (RoQ) has been identified which is very much difficult to detect. The RoQ attacks can use source and destination IP address spoofing, and they do not have distinct periodicity, and may not filter the attack packets precisely. In this paper, we propose to design the MAC layer based defense architecture for RoQ attacks in Wireless LAN which includes the detection and response stages. The attackers are detected by checking the RTS/CTS packets from the MAC layer and the corresponding attack flows are blocked or rejected. By our simulation results, we show that our proposed technique achieves reduces the attack throughput there by increasing the received bandwidth and reducing the packet loss of legitimate users.

1. Introduction

The characteristic vulnerabilities of LAN (802.11x) networks makes the attack possible. Like wired networks, it is not possible to physically protect the wireless networks. The next offices, the parking lot of the building, across the street or possibly several miles away, are the source of attacks for the wireless networks, i.e. the attack can be carried out from anyplace. It is necessary to realize the facts of different attacks in opposition to the wireless infrastructure to establish the appropriate defense strategy. The execution of several attacks which are not dangerous can be performed easily while the attacks which have destructive effects can be seriously complicated. The risk which is included in the wireless security is more when compared with any other case of security.

There are many potentially disturbing threats to wireless local area networks (WLANs). The security issues which are ranging from misconfigured wireless access points (WAPs) to session hijacking to Denial of Service (DoS) affects the WLAN. Wireless networks are also vulnerable on the particular threats in the wide array of 802.11, other than the TCP/IP based attacks which are related to the wired networks. A security solution which includes an intrusion detection system (IDS) should be used by the WLANs in order to assist in the defense and detection of these possible threats. Even organizations not having a WLAN should consider an IDS solution since it may be dangerous due to wireless threats [2].

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

WLAN are exposed to a variety of threats. The standard 802.11 encryption method, known as Wired Equivalent Privacy (WEP), is weaker.

The hackers attack a WLAN and collect the sensitive data by introducing a misbehaving WAP into the WLAN coverage area. The misbehaving WAP can be designed like an actual WAP because several wireless clients are connected simply to the WAP with the best signal strength. Moreover, the users can be "trapped" to connect with the misbehaving WAP unintentionally. When a user is associated, all the communications can be monitored by the hacker through the misbehaving WAP. Apart form the hackers, misbehaving nodes can also be introduced by the users. When low cost and easy implementation is combined with the flexibility of the wireless network communications, the WLANs can be very attractive to the users. By installing a WAP on a recognized LAN, the users can create a backdoor into the network, undermining all the hardwired security solutions and give the network open to the hackers [2].

Since the networks which are using 802.11 are vulnerable to numerous Denial of service (DoS) attacks, WLAN can be made fatal. Wireless communications which are uncertain upon physical objects are naturally vulnerable to signal degradation. Also, the hackers can overflow WAPs with association requests and injects the malicious DoS attacks there by forcing them to reboot. Moreover, by sending a repeated disassociate/deauthenticate requests with the help of the above mentioned rouge WAP, the hackers can refuse service to a wireless client.

Still there are various threats to WLAN and the identification of further vulnerabilities is performed at a high speed. The general truths are the reality of the threats, their ability to create broad destruction and their rising familiarity with increase in fame of the 802.11 technology. With the absence of the detection mechanism, the identification of the threats to a WLAN can be complicated. When the consciousness of the threats is absent then a network is inadequately secure with respect to the threats facing it. When the threats to the networks are recognized, then it is possible to equip the WLAN suitably with the necessary security measures.

1.1 RoQ Attack

By the high rate or high volume, the typical DDoS flooding attacks are characterized. An alternative of DDoS attacks has been identified recently which is too complex to detect which are called as shrew attacks or Reduction-of-Quality (RoQ) attacks. Instead of refusing the clients from the services completely, these RoQ attacks throttle the TCP throughput heavily and reduce the QoS to end systems gradually [3].

The transients of systems adaptive behavior is targeted by the RoQ attacks instead of limiting its steady-state capacity. The RoQ attacks can use source and destination IP address spoofing, and they do not have distinct periodicity, and may not filter the attack packets precisely. In order to escape from being caught by the traceback techniques, RoQ attacks often launch attacks through multiple zombies and spoof header packet information. But, it is important to control the frequency domain characteristics of attacking flows. In order to throttle the TCP flows efficiently, the attacking period has to be close to the Retransmission Time Out (RTO). Using traffic spectrum, the energy distribution pattern will give up such malicious flow detection mechanisms even if the source IP addresses are carried in packet header are falsified [3].

In this paper, we propose to design a MAC layer based defense architecture for Reduction-of-Quality (RoQ) attacks in Wireless LAN. It includes the detection and isolation of attackers. Detection makes use of three status values that can be obtained from the MAC layer: Frequency of receiving RTS/CTS packets, frequency of sensing a busy channel and the number of RTS/DATA retransmissions. Once the attackers are detected, the corresponding flows are blocked or rejected.

2. Related Work

Yu Chen et al [3] have explored the energy distributions of Internet traffic flows in frequency domain. Normal TCP traffic flow

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

present periodicity because of protocol behavior. Their results revealed that normal TCP flows can be segregated from malicious flows according to energy distribution properties. They have discovered the spectral shifting of attack flows from that of normal flows. Combining flow-level spectral analysis with sequential hypothesis testing, they have proposed a novel defense scheme against RoQ attacks. Their detection and filtering scheme can effectively rescue 99% legitimate TCP flows under the RoQ attacks.

Zhongua Zhang et al [4] focused on the examination of the anomaly-based intrusion detector's operational capabilities and drawbacks through their operating environments. Anomaly detection is classified in a statistical framework based on the similarity with the induction problem for describing their general expected behaviors. For the apparent subjects from hosts and networks, several key problems and respective potential solutions about the normality characterization for the observable subjects are addressed. Based on some existing achievements anomaly detector's evaluations are also examined.

Piyush Kumar Shukla et al [5] have analyzed the congestion based DDoS attacking in mobile ad hoc networks. They have proposed a grammar based approach to modeling and analyzing multi-step network attack sequences. Moreover, they have proposed a low rate DoS attack detection algorithm, which relies on the core characteristic of the low rate DoS attack in introducing high rate traffic for short periods, and then uses a proactive test based differentiation technique to filter the attack packets. They have evaluated the feasibility of the proposed low rate DoS attack algorithm on real Internet traces.

Martin Rehak et al [6] have proposed a research to detect malicious traffic in high-speed networks by correlated anomaly detection methods. Based on FPGA elements transparent inline probes are used to obtain the real time traffic statistics in NetFlow format and gives a traffic statistics to the agent-based detection layer. The agent uses a particular anomaly detection method in this layer to detect the anomalies and describes the flows in its extended trust model. The agent shares the anomaly estimation of the individual network flows which are uses as an input for the agents trusts models. In order to estimate their maliciousness the trustfulness values of individual flows from all agents are combined.

John Haggerty et al [7] have proposed that a major threat to the information economy is denial-of-service attacks. These attacks are common even though the widespread usage of the perimeter model countermeasures. Therefore to provide early detection of flooding denial-of-service attacks, a new approach is assumed which uses statistical signatures at the router. There are three advantages for this approach. They are: Computational load on the defense mechanism is reduced by analyzing fewer packets, if the system is under protection then the state information is not required and alerts may span many attack packets. Thus to prevent malicious packets from reaching their proposed target in the first the defense mechanism may be placed within the routing infrastructure.

Mina Guirguis et al [8] have analytically captured the effect of RoQ attacks that would deprive an Internet element from reaching steady state by knocking it off whenever it is about to stabilize. They have formalized the notion of attack "potency", which exposes the tradeoff between the "damage" inflicted by an attacker and the "cost" of the attack. Moreover, their notion takes aggressiveness into account which enabled them to identify different families of DoS attacks based on their aggressiveness.

Mina Guirguis et al [9] have exemplified the security implications by exposing the vulnerabilities of admission control mechanisms that are widely deployed in Internet end systems to Reduction of Quality (RoQ) attacks. They have shown that a well orchestrated RoQ attack on an end-system admission control policy could introduce significant inefficiencies that could potentially deprive an Internet end-system from much of its capacity, or significantly reduce its service quality, while evading detection by consuming an unsuspicious, small fraction of that system's hijacked capacity. They have developed a control theoretic model for assessing the impact of RoQ attacks on an end-system's admission controller. They quantified the damage inflicted by an attacker through deriving appropriate metrics.

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

Eduardo Mosqueira-Rey et al [10] have described the design of misuse detection agent which is one of the different agents in a multiagent-based intrusion detection system. Using a packet sniffer the agent examines the packets in the network connections and creates a data model based on the information obtained. This data model is the input to the rule based agent inference engine which uses the Rete algorithm for pattern matching. So the rules of the signature-based intrusion detection system become small.

Magnus Almgren et al [11] have investigated the procedure to use the alerts from may audit sources to improve the accuracy of the intrusion detection system (IDS). A theoretical model is designed automatically for the reason about the alerts from the different sensors through concentrating on the web server attacks. It also provides a better understanding of possible attacks against their systems for the security operators. This model enables reasoning about the absence of the expected alerts by taking the sensor status and its capability into account. This model is built using Bayesian networks which needs some initial parameter values that can be provided from the IDS operator.

Naeimeh Laleh et al [12] have proposed that fraud is growing remarkably with the growth of modern technology and the universal superhighways of communication which results in the loss of billions of dollars throughout the world each year. This technique tends to propose a new taxonomy and complete review for the different types of fraud and data mining techniques of fraud detection. The uniqueness of this technique is gathering all types of frauds which can be detected by data mining techniques and analyzes some real time approaches which have the ability to detect the frauds in real time.

Yu Chen et al [13] have proposed a new signal-processing approach to identify and detect the attacks by examining the frequency domain characteristics of incoming traffic flows to a server. Their proposed technique is effective in that its detection time is less than a few seconds. Furthermore, their technique entails simple implementation, making it deployable in real-life network environments.

3. Proposed Defense Technique

In this paper, we propose a defense scheme that includes the detection and response stages.

Detection makes use of three status values that can be obtained from the MAC layer: frequency of receiving RTS/CTS packets, frequency of sensing a busy channel, and the number of RTS/DATA retransmissions. When the number of RTS/CTS packets received exceeds a certain threshold RC_{th}, it indicates that too many nodes are within the transmission range to compete for the channel. When the channel is sensed to be in a busy state, a node will persist in the backoff stage and stop the CW count. When the stopping time is longer than a threshold SE_{th}, it indicates that too many nodes are within the interference range. Thus if the number of retransmissions exceeds a threshold RE_{th}, it will be regarded as an indicator for channel congestion. Since these status values are already available in the protocol stack implementation, the overhead required for implementing this detection scheme is very low.

During the response phase, the nodes will check the following conditions to mark each packet with a Congestion Bit (CB)

1. If number of RTS/CTS packets > RC_{th},
2. If Stime > SE_{th},
3. If number of RTS/DATA retransmissions > RE_{th},

Now, the congestion bit is set as follows:

CB = 000: If none of the Conditions are true.

CB = 100: If Condition 1 is true.

CB = 010: If Condition 2 is true.

CB = 001: If Condition 3 is true.

CB = 110: If Conditions 1&2 are true.

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

CB = 101: If Conditions1&3 are true.

CB = 011: If Conditions 2&3 are true.

CB = 111: If all the Conditions are true.

We propose that RTS and CTS values between two communicating nodes can be observed by a passive server and it calculates the CB. This helps to detect and isolate the attackers.

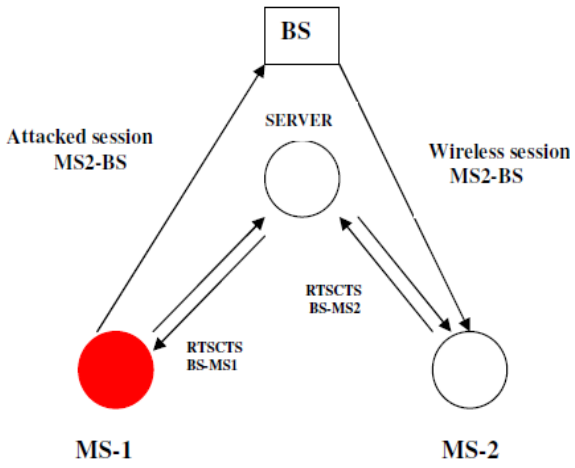


Fig. 1 RTS/CTS monitoring by Server

The Server executes the following algorithm, to detect the attackers.

Algorithm

1. Let initial time interval = t1

2. Server checks RTS/CTS packets and calculates CB.

3. If CB ≠ 000, then

3.1 If CB = 100 or 010 or 001, then

3.1.1 Mark the source status as normal and transmit CB to the source.

3.2 Else if CB = 110 or 101 or 011, then

3.2.1 Mark the source status as suspected and transmit CB to the source.

3.3 Else if CB = 111, then

3.3.1 Mark the source status as attacker and transmit CB to the source.

3.4 End if

4. End if

5. t1 = t1 + 1

6. Repeat the steps from 2

7. If t1 = 3 and source status = attacker, then

6.1 Remove the node from the list.

6.2 Block all the traffic from the attacker

8. Else If t1 = 4 and source status = suspected, then

8.1 Remove the node from the list.

8.2 Block all the traffic from the attacker

9. End if

At the initial time interval t1, the server checks the RTS/CTS packets and calculates the Congestion Bit (CB) according to the conditions mentioned above. If the value of CB is either 100 or 010 or 001, then the status of the source is marked as normal and the CB is transmitted to the source. If the value of CB is either 110 or 101 or 011, then the status of the source is marked as suspected and the CB is transmitted to the source. If the value of CB is 111 the status of the source is marked as attacker and transmits the CB to the source. Suppose, if the status of the source is still attacker till the time interval t3, then the corresponding node is removed from the list and all the traffics from the attackers are blocked. Similarly, suppose if the status of the source is still suspected till the time interval t4, then the corresponding node is removed from the list and all the traffics from the attackers are blocked.

4. Simulation Results

4.1 Simulation Model and Parameters

This section deals with the experimental performance evaluation of our algorithm through

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

simulations. In order to test our protocol, the NS2 [14] simulator is used. We compare our proposed MAC Layer Based Defense Architecture for Reduction-of-Quality (RoQ) with Shrew [13] filter.

4.2 Performance Metrics

In our experiments, we measure the following metrics

- Received Bandwidth
- Packet Loss

The simulation results are described in the next section.

4.3 Results

A. Effect of Varying Attackers

In our first experiment, we vary the number of attackers as 2, 4, 6 and 8 in order to calculate the received bandwidth and packet loss legitimate users.

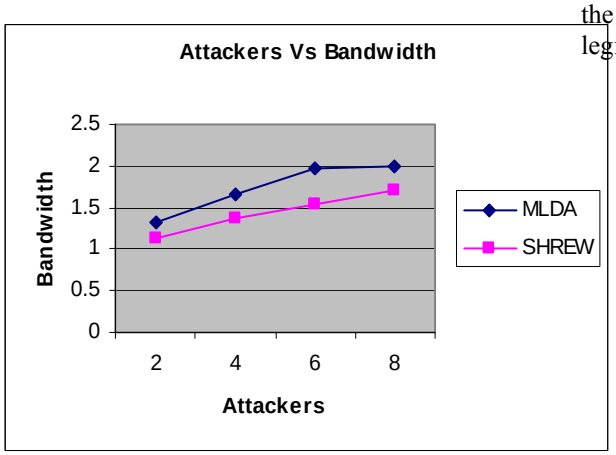


Fig: 2 Attackers Vs Bandwidth

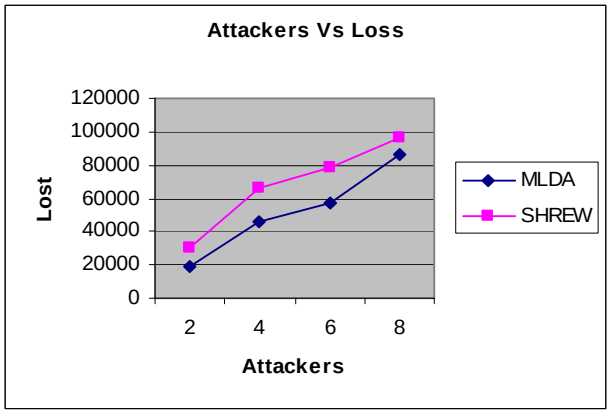


Fig: 3 Attackers Vs Loss

Fig: 2 gives the received bandwidth for normal legitimate users when varying the number of attackers. It shows that the bandwidth received for normal users is more in the case of MLDA when compared with SHREW.

Fig: 3 illustrates that the packet loss due to attack is more in SHREW when compared with MLDA, when varying the number of attackers.

.

B. Effect of Varying Attack Period

In our final experiment, we vary the number of attack period as 0, 5…20 in order to calculate the received bandwidth and packet loss of the legitimate users.

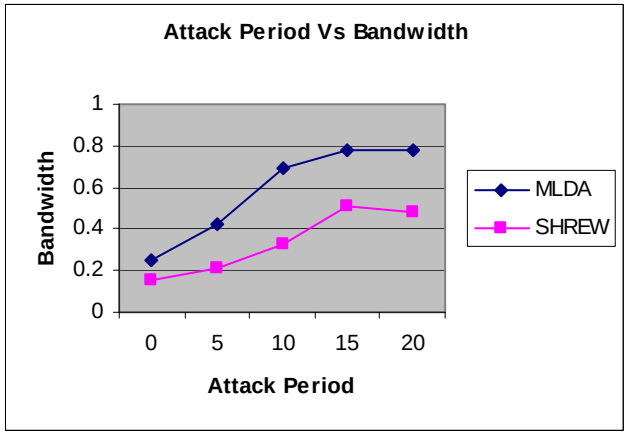


Fig: 4 Attack Period Vs Bandwidth

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

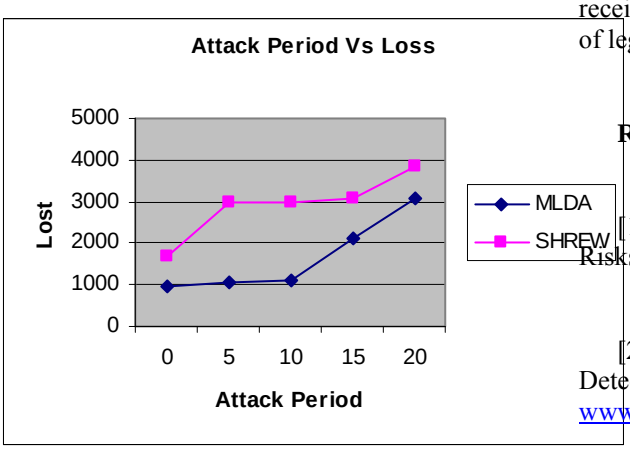


Fig: 5 Attack Period Vs Loss

Fig: 4 gives the received bandwidth for normal legitimate users when varying the number of attack period. It shows that the bandwidth received is more in the case of MLDA when compared with SHREW.

Fig: 5 illustrates that the packet loss due to attackers is more in SHREW when compared with MLDA.

5. Conclusion

In this paper, to defend against the Reduction-of-Quality (RoQ) attacks, we have proposed a MAC layer based defense architecture in Wireless LAN. It includes the detection and response stages. Detection makes use of three status values that can be obtained from the MAC layer: frequency of receiving RTS/CTS packets, frequency of sensing a busy channel, and the number of RTS/DATA retransmissions. In our proposed architecture, RTS and CTS values between two communicating nodes can be observed by a passive server to detect and isolate the attackers. It calculates a cumulative congestion bit (CB), depending on these three status values. By executing our algorithm the server detects the attackers. Once the attackers are detected, the corresponding attack flows are blocked or rejected. By our simulation results, we have shown that our proposed technique reduces the attack throughput there by increasing the received bandwidth and reducing the packet loss of legitimate users.

References

[1] Kimmo Hiltunen, "WLAN Attacks and Risks", White Paper, Ericson, January 2008.

[2] Jamil Farshchi, "Wireless Intrusion Detection System", Security Focus, Nov- 2003, www.securityfocus.com/infocus/1742

[3] Yu Chen and Kai Hwang, "Spectral Analysis of TCP Flows for Defense against Reduction-of-Quality Attacks", in the IEEE International Conference on Communications-ICC-2007, June 2007.

[4] Zonghua Zhang and Hong Shen, "A Brief Observation Centric Analysis on Anomaly Based Intrusion Detection", Springer-Verlag Berlin, Heidelberg 2005.

[5] Piyush Kumar Shukla, S. Silakari and S.S. Bhadouria, "Designing And Analysis Issues For An Attack Resilient and Adaptive Medium Access Control Protocol for Computer Networks: An Exclusive Survey",

[6] Martin Rehak, Michal pechoucek, karel Bartos, Martin Grill, Pavel celeda and vojtech krmick "An intrusion detection system for high-speed networks", national institute of informatics, 2008.

[7] John Haggerty, Qi Shi and Madjid Merabti, "Statistical Signatures For Early Detection Of Flooding Denial-Ofservice Attacks", Springer Boston, 2006.

[8] Mina Guirguis, Azer Bestavros and Ibrahim Matta, "Exploiting the Transients of Adaptation for RoQ Attacks on Internet Resources", 12th IEEE International Conference on Network Protocols (ICNP'04).

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

[9] Mina Guirguis, Azer Bestavros, Ibrahim Matta and Yuting Zhang, "Reduction of Quality (RoQ) Attacks on Internet End-Systems", Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2005.

[10] Eduardo Mosqueira-Rey, Amparo Alonso-Betanzos, Belen Baldonedo Del Rio, and Jesus Lago Pineiro, "A Misuse Detection Agent for Intrusion Detection in a Multi-agent Architecture", Springer-Verlag, Berlin Heidelberg 2007.

[11] Magnus Almgren, Ulf Lindqvist, and Erland Jonsson, "A Multi-Sensor Model to Improve Automated Attack Detection", Springer-Verlag Berlin Heidelberg 2008.

[12] Naeimeh Laleh and Mohammad Abdollahi Azgomi, "A Taxonomy of Frauds and Fraud Detection Techniques", Springer-Verlag Berlin Heidelberg 2009.

[13] Yu Chen, Yu-Kwong Kwok, and Kai Hwang, "Filtering Shrew DDoS Attacks Using A New Frequency-Domain Approach", In the IEEE Conference on Local Computer Networks, 2005, 30th Anniversary.

[14] Network Simulator: www.isi.edu/ns

About the Authors

Jatinder Singh *Received his M.Tech degree from Punjabi University, Patiala in 2003.He is a dynamic researcher and prolific author in the field of Computer Engineering. He has also won Best Research Scholal Award by UGC and Management Excellence Award by MIDI, Punjab. He published 50 National and International research papers over the years as well as 20 highly acclaimed text and research books. He is also a member of several professional scientific organizations and has lectured widely at academic institutions in India and Abroad.Presently is working as a Director in Universal Institute of Engg. & Tech. Lalru CHD.*

Lakhwinder Kaur *received the M.E. degree from TIET, Patiala, Punjab, in 2000 and Ph.D. degree from PTU, Jalandhar in 2007 both in computer science and engineering . She has been in the teaching profession since 1992. Presently, she is working as Reader in the Department of CSE, University College of Engg., Punjabi University, Patiala (Pb). Her research interests include image compression and denoising, Grid computing and wavelets.*

Savita Gupta *received the B.Tech. degree from TITS, Bhiwani (Haryana), in 1992, M.E. degree from TIET, Patiala, Punjab, in 1998 both in computer science and engineering. She obtained her Ph.D. degree from PTU, Jalandhar in 2007 in the field of Ultrasound Image Processing. She has been in the teaching profession since 1992. Presently, she is working as Professor in the Department of CSE, University Instiute of of Engg. & Technology, Panjab University, Chandigarh. Her research interests include image processing, image compression and denoising, and wavelet applications*

---

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

Application of k-Means Clustering algorithm for prediction of Students' Academic Performance

Oyelade, O. J
Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria.
Ola2000faith@yahoo.co.uk.

Oladipupo, O. O
Department of Computer and Information Sciences, College of Science and Technology, Covenant University, Ota, Nigeria.
frajooye@yahoo.com.

Obagbuwa, I. C
Department of Computer Science
Lagos State University, Lagos, Nigeria.
ibidunobagbuwa@yahoo.com

***Abstract*— The ability to monitor the progress of students' academic performance is a critical issue to the academic community of higher learning. A system for analyzing students' results based on cluster analysis and uses standard statistical algorithms to arrange their scores data according to the level of their performance is described. In this paper, we also implemented k-mean clustering algorithm for analyzing students' result data. The model was combined with the deterministic model to analyze the students' results of a private Institution in Nigeria which is a good benchmark to monitor the progression of academic performance of students in higher Institution for the purpose of making an effective decision by the academic planners.**

Keywords- k – mean, clustering, academic performance, algorithm.

I. INTRODUCTION

Graded Point Average (GPA) is a commonly used indicator of academic performance. Many Universities set a minimum GPA that should be maintained in order to continue in the degree program. In some University, the minimum GPA requirement set for the students is 1.5. Nonetheless, for any graduate program, a GPA of 3.0 and above is considered an indicator of good academic performance. Therefore, GPA still remains the most common factor used by the academic planners to evaluate progression in an academic environment [1]. Many factors could act as barriers to students attaining and maintaining a high GPA that reflects their overall academic performance during their tenure in University. These factors could be targeted by the faculty members in developing strategies to improve student learning and improve their academic performance by way of monitoring the progression of their performance.

Therefore, performance evaluation is one of the bases to monitor the progression of student performance in higher Institution of learning. Base on this critical issue, grouping of students into different categories according to their performance has become a complicated task. With traditional grouping of students based on their average scores, it is difficult to obtain a comprehensive view of the state of the students' performance and simultaneously discover important details from their time to time performance.

With the help of data mining methods, such as clustering algorithm, it is possible to discover the key characteristics from the students' performance and possibly use those characteristics for future prediction. There have been some promising results from applying k-means clustering algorithm with the Euclidean distance measure, where the distance is computed by finding the square of the distance between each scores, summing the squares and finding the square root of the sum [6].

This paper presents k-means clustering algorithm as a simple and efficient tool to monitor the progression of students' performance in higher institution.

Cluster analysis could be divided into hierarchical clustering and non-hierarchical clustering techniques. Examples of hierarchical techniques are single linkage, complete linkage, average linkage, median, and Ward. Non-hierarchical techniques include k-means, adaptive k-means, k-medoids, and fuzzy clustering. To determine which algorithm is good is a function of the type of data available and the particular purpose of analysis. In more objective way, the stability of clusters can be investigated in simulation studies [4]. The problem of selecting the "best" algorithm/parameter setting is a difficult one. A good clustering algorithm ideally should produce groups with distinct non-overlapping boundaries, although a perfect separation can not typically be achieved in practice. Figure of merit measures (indices) such as the silhouette width [4] or the homogeneity index [5] can be used to evaluate the quality of separation obtained using a clustering algorithm. The concept of stability of a clustering algorithm was considered in [3]. The idea behind this validation approach is that an algorithm should be rewarded for consistency. In this paper, we implemented traditional k-means clustering algorithm [6] and Euclidean distance measure of similarity was chosen to be used in the analysis of the students' scores.

II. METHODOLOGY

A. Development of k-mean clustering algorithm

Given a dataset of n data points $x_1, x_2, \ldots, x_n$ such that each data point is in $\mathbf{R}^d$, the problem of finding the minimum

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

variance clustering of the dataset into k clusters is that of finding k points $\{m_j\}$ (j=1, 2, …, k) in $\mathbf{R}^d$ such that

$$\frac{1}{n}\sum_{i=1}^{n}\left[\min_j d^2(x_i,m_j)\right] \quad (1)$$

is minimized, where $d(x_i, m_j)$ denotes the Euclidean distance between x_i and m_j. The points $\{m_j\}$ (j=1, 2, …,k) are known as cluster centroids. The problem in Eq.(1) is to find k cluster centroids, such that the average squared Euclidean distance (mean squared error, MSE) between a data point and its nearest cluster centroid is minimized.

The k-means algorithm provides an easy method to implement approximate solution to Eq.(1). The reasons for the popularity of k-means are ease and simplicity of implementation, scalability, speed of convergence and adaptability to sparse data.

The k-means algorithm can be thought of as a gradient descent procedure, which begins at starting cluster centroids, and iteratively updates these centroids to decrease the objective function in Eq.(1). The k-means always converge to a local minimum. The particular local minimum found depends on the starting cluster centroids. The problem of finding the global minimum is NP-complete. The k-means algorithm updates cluster centroids till local minimum is found. Fig.1 shows the generalized pseudocodes of k-means algorithm; and traditional k-means algorithm is presented in fig. 2 respectively.

Before the k-means algorithm converges, distance and centroid calculations are done while loops are executed a number of times, say l, where the positive integer l is known as the number of k-means iterations. The precise value of l varies depending on the initial starting cluster centroids even on the same dataset. So the computational time complexity of the algorithm is $O(nkl)$, where n is the total number of objects in the dataset, k is the required number of clusters we identified and l is the number of iterations, $k \leq n$, $l \leq n$ [6].

Step 1: Accept the number of clusters to group data into and the dataset to cluster as input values

Step 2: Initialize the first K clusters
- Take first k instances or
- Take Random sampling of k elements

Step 3: Calculate the arithmetic means of each cluster formed in the dataset.

Step 4: K-means assigns each record in the dataset to only one of the initial clusters
- Each record is assigned to the nearest cluster using a measure of distance (e.g Euclidean distance).

Step 5: K-means re-assigns each record in the dataset to the most similar cluster and re-calculates the arithmetic mean of all the clusters in the dataset.

Fig 1: Generalised Pseudocode of Traditional k-means

```
1   MSE = largenumber;
2   Select initial cluster centroids {m_j}_j
    K = 1;
3   Do
4     OldMSE = MSE;
5     MSE1 = 0;
6     For j = 1 to k
7       m_j = 0; n_j = 0;
8     endfor
9     For i = 1 to n
10      For j = 1 to k
11      Compute squared Euclidean
        distance d^2(x_i, m_j);
12      endfor
13    Find the closest centroid m_j to x_i;
14      m_j = m_j + x_i; n_j = n_j+1;
15      MSE1=MSE1+ d^2(x_i, m_j);
16    endfor
17    For j = 1 to k
18      n_j = max(n_j, 1); m_j = m_j/n_j;
19    endfor
20  MSE=MSE1;
    while (MSE<OldMSE)
```

Fig.2: Traditional k-means algorithm [6]

III. RESULTS

We applied the model on the data set (academic result of one semester) of a university in Nigeria. The result generated is shown in tables 2, 3, and 4, respectively. In table 2, for k = 3; in cluster 1, the cluster size is 25 and the overall performance is 62.22. Also, the cluster sizes and the overall performances for cluster numbers 2 and 3 are 15, 29 and 45.73, and 53.03, respectfully. Similar analyses also hold for tables 3 and 4. The graphs are generated in figures 3, 4 and 5, respectively, where the overall performance is plotted against the cluster size.

Table 5 shows the dimension of the data set (Student's scores) in the form N by M matrices, where N is the rows (# of students) and M is the column (# of courses) offered by each student.

The overall performance is evaluated by applying deterministic model in Eq. 2 [7] where the group assessment in each of the cluster size is evaluated by summing the average of the individual scores in each cluster.

$$\frac{1}{N}\left(\sum_{j=1}^{N}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)\right) \quad 2$$

Where

N = the total number of students in a cluster and
n = the dimension of the data

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

Table 1: Performance index

70 and above	Excellent
60-69	Very Good
50-59	Good
45-49	Very Fair
40-45	Fair
Below 45	Poor

In Figure 3, the overall performance for cluster size 25 is 62.22% while the overall performance for cluster size 15 is 45.73% and cluster size 29 has the overall performance of 53.03%. This analysis showed that, 25 out of 79 students had a "Very Good" performance (62.22%), while 15 out of 79 students had performance in the region of very "Fair" performance (45.73%) and the remaining 29 students had a "Good" performance (53.03%) as depicted in the performance index in table 1.

Figure 4 shows the trends in performance analysis as follows; overall performance for cluster size 24 is 50.08% while the overall performance for cluster size 16 is 65.00%. Cluster size 30 has the overall performance of 58.89%, while cluster size 09 is 43.65%. The trends in this analysis indicated that, 24 students fall in the region of "Good" performance index in table 1 above (50.08%), while 16 students has performance in the region of "Very Good" performance (65.00%). 30 students has a "Good" performance (58.89%) and 9 students had performance of "Fair" result (43.65%).

In figure 5, the overall performance for cluster size 19 is 49.85%, while the overall performance for cluster size 17 is 60.97%. Cluster size 9 has the overall performance of 43.65%, while the cluster size 14 has overall performance of 64.93% and cluster size 20 has overall performance of 55.79%. This performance analysis indicated that, 19 students crossed over to "Good" performance region (49.85%), while 17 students had "Very Good" performance results (60.97%). 9 students fall in the region of "Fair" performance index (43.65%), 14 students were in the region of "Very Good" performance (64.93%) and the remaining 20 students had "Good" performance (55.79%).

Table 2: K = 3

Cluster #	Cluster size	Overall Performance
1	25	62.22
2	15	45.73
3	29	53.03

Table 3: K = 4

Cluster #	Cluster size	Overall Performance
1	24	50.08
2	16	65.00
3	30	58.89
4	9	43.65

Table 4: K = 5

Cluster #	Cluster size	Overall Performance
1	19	49.85
2	17	60.97
3	9	43.65
4	14	64.93
5	20	55.79

Table 5: Statistics of the Data used

Student's Scores	Number of Students	Dimension(Total number of courses)
Data	79	9

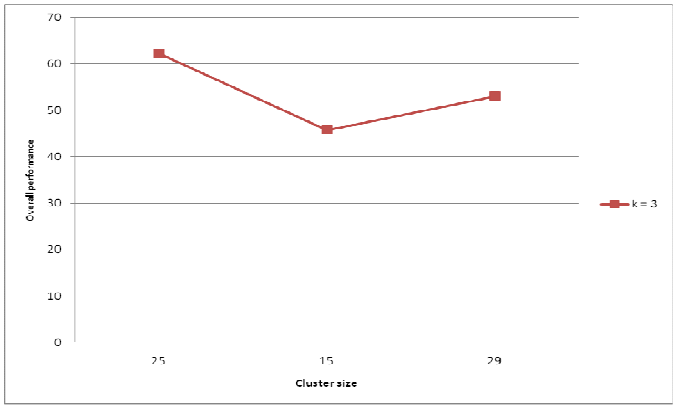


Fig. 3: Overall Performance versus cluster size (# of students) k = 3

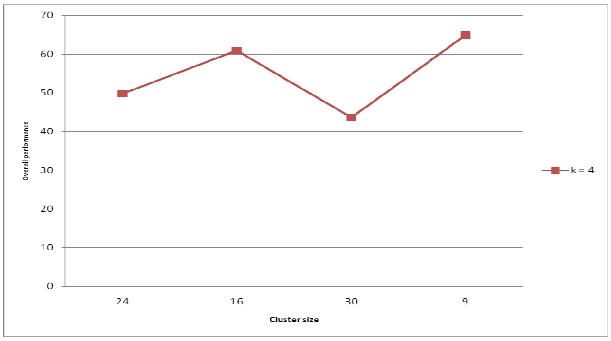


Fig. 4: Overall Performance versus cluster size (# of students) k = 4

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

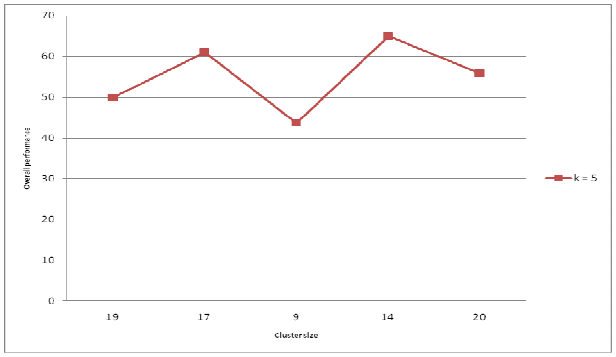


Fig. 5: Overall Performance versus cluster size (# of students) k = 5

IV. DISCUSSION AND CONCLUSION

In this paper, we provided a simple and qualitative methodology to compare the predictive power of clustering algorithm and the Euclidean distance as a measure of similarity distance. We demonstrated our technique using k-means clustering algorithm [6] and combined with the deterministic model in [7] on a data set of private school results with nine courses offered for that semester for each student for total number of 79 students, and produces the numerical interpretation of the results for the performance evaluation. This model improved on some of the limitations of the existing methods, such as model developed by [7] and [8]. These models applied fuzzy model to predict students' academic performance on two dataset only (English Language and Mathematics) of Secondary Schools results. Also the research work by [9] only provides Data Mining framework for Students' academic performance. The research by [10] used rough Set theory as a classification approach to analyze student data where the Rosetta toolkit was used to evaluate the student data to describe different dependencies between the attributes and the student status where the discovered patterns are explained in plain English.

Therefore, this clustering algorithm serves as a good benchmark to monitor the progression of students' performance in higher institution. It also enhances the decision making by academic planners to monitor the candidates' performance semester by semester by improving on the future academic results in the subsequence academic session.

ACKNOWLEDGMENT

This work was funded by Covenant University Center for Research and Development. We are grateful to Fahim A. M. and Salem A. M. for their useful materials. We also thank Dr. Obembe for his useful assessment which has improved on the quality of this work.

REFERENCES

[1] S. Sujit Sansgiry, M. Bhosle, and K. Sail, "Factors that affect academic performance among pharmacy students," American Journal of Pharmaceutical Education, 2006.

[2] Susmita Datta and Somnath Datta, "Comparisons and validation of statistical clustering techniques for microarray gene expression data," Bioinformatics, vol. 19, pp.459–466, 2003.

[3] Rousseeuw P. J, "A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational Appl Math, vol 20, pp. 53–65, 1987.

[4] Sharmir R. and Sharan R., "Algorithmic approaches to clustering gene expression data," In current Topics in Computational Molecular Biology MIT Press; pp. 53-65, 2002.

[5] Mucha H. J., "Adaptive cluster analysis, classification and multivarite graphics,"Weirstrass Institute for Applied Analysis and Stochastics, 1992.

[6] Fahim A. M., Salem A. M., Torkey F. A. and Ramadan M. A., "An efficient enhanced k-means clustering algorithm," Journal of Zhejiang University Science A., pp. 1626–1633, 2006

[7] J. O. Omolehin, J. O. Oyelade, O. O. Ojeniyi and K. Rauf, "Application of Fuzzy logic in decision making on students' academic performance," Bulletin of Pure and Applied Sciences, vol. 24E(2), pp. 281-187, 2005.

[8] J. O. Omolehin, A. O. Enikuomehin, R. G. Jimoh and K. Rauf, "Profile of conjugate gradient method algorithm on the performance appraisal for a fuzzy system," African Journal of Mathematics and Computer Science Research," vol. 2(3), pp. 030-037, 2009.

[9] N. V. Anand Kumar and G. V. Uma, "Improving Academic Performance of Students by Applying Data Mining Technique," European Journal of Scientific Research, vol. 34(4), 2009.

[10] Varapron P. et al., "Using Rough Set theory for Automatic Data Analysis," 29th Congress on Science and Technology of Thailand, 2003.

AUTHORS PROFILE

Oyelade, O. J.: Received his Bachelor degree in Computer Science with Mathematics(Combined Honour) and M.Sc. in Computer Science from Obafemi Awolowo University, Ile-Ife, Nigeria. He is a Ph.D. Candidate, and a Faculty member in the Department of Computer and Information Sciences, Covenant University, Nigeria. His research interests are in Bioinformatics, Clustering, Fuzzy logic and Algorithms.

Oladipupo, O.O.: Received her Bachelor degree in Computer Science from University of Ilorin, and M.Sc. in Computer Science from Obafemi Awolowo University, Ile-Ife, Nigeria. She is a Ph.D. Candidate, and a Faculty member in the Department of Computer and Information Sciences, Covenant University, Nigeria. Her research interests are in Artificial Intelligent, Data mining and Soft Computing Techniques.

Obagbuwa, I. C.: Received her Bachelor degree in Computer Science from University of Ilorin, Ilorin, Nigeria and Master degree (M.Sc.) in Computer Science from University of Port Harcourt, Port Harcourt, Nigeria. She is a Ph.D Candidate in University of Port Harcourt, Port Harcourt, Nigeria in Computer Science. She is a Faculty member in the Department of Computer Sciences, Lagos State University, Ojo – Lagos, Nigeria. Her research interests are in Text segmentation/Automatic Information Extraction, Databases, Document management, Telecommunication and Networking.

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

IJCSIS REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA

Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia

Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA

Dr. Emanuele Goldoni, University of Pavia, Dept. of Electronics, Italy

Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway

Assoc. Prof. N. Jaisankar, VIT University, Vellore,Tamilnadu, India

Dr. Amogh Kavimandan, The Mathworks Inc., USA

Dr. Ramasamy Mariappan, Vinayaka Missions University, India

Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China

Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA

Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico

Dr. Neeraj Kumar, SMVD University, Katra (J&K), India

Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania

Dr. Junjie Peng, Shanghai University, P. R. China

Dr. Ilhem LENGLIZ, HANA Group - CRISTAL Laboratory, Tunisia

Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India

Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain

Prof. Dr.C.Suresh Gnana Dhas, Anna University, India

Mrs Li Fang, Nanyang Technological University, Singapore

Prof. Pijush Biswas, RCC Institute of Information Technology, India

Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia

Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India

Mr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand

Mr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India

Mr. Hayder N. Jasem, University Putra Malaysia, Malaysia

Mr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India

Mr. R. S. Karthik, C. M. S. College of Science and Commerce, India

Mr. P. Vasant, University Technology Petronas, Malaysia

Mr. Wong Kok Seng, Soongsil University, Seoul, South Korea

Mr. Praveen Ranjan Srivastava, BITS PILANI, India

Mr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong

Mr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia

Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan

Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria

Dr. Riktesh Srivastava, Skyline University, UAE

Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia

Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt

CALL FOR PAPERS
International Journal of Computer Science and Information Security
IJCSIS 2010
ISSN: 1947-5500
http://sites.google.com/site/ijcsis/

International Journal Computer Science and Information Security, now at its sixth edition, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2009-2010 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span
a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc,), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on
its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embeded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://sites.google.com/site/ijcsis/authors-notes .

ISSN 1947 5500