Online ISSN : 0975-4172
Print ISSN : 0975-4350

Global Journal of Computer Science & Technology

DISCOVERING THOUGHTS AND INVENTING FUTURE

Highlights

- E-Voting System
- Concept of Histogram
- Salt & Pepper Noise
- Handling of Congestion

Laboratory, Antarctica

Volume 12 | Issue 5 | Version 1.0 ENG

GLOBAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY

Global Journal of Computer Science and Technology

Volume 12 Issue 5 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

(A Delaware USA Incorporation with "Good Standing"; Reg. Number: 0423089)

Sponsors:Global Association of Research

Open Scientific Standards

Publisher's Headquarters office

Global Journals Inc., Headquarters Corporate Office,
Cambridge Office Center, II Canal Park, Floor No. 5th, ***Cambridge (Massachusetts)***, Pin: MA 02141
United States
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Association of Research, Marsh Road, Rainham, Essex, London RM13 8EU
United Kingdom.

Packaging & Continental Dispatching

Global Journals, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investers@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Including by Air Parcel Charges):

For Authors:
22 USD (B/W) & 50 USD (Color)
Yearly Subscription (Personal & Institutional):
200 USD (B/W) & 250 USD (Color)

Editorial Board Members (HON.)

Dr. Bart Lambrecht
Director of Research in Accounting and FinanceProfessor of Finance
Lancaster University Management School
BA (Antwerp); MPhil, MA, PhD (Cambridge)

Dr. Carlos García Pont
Associate Professor of Marketing
IESE Business School, University of Navarra
Doctor of Philosophy (Management), Massachusetts Institute of Technology (MIT)
Master in Business Administration, IESE, University of Navarra
Degree in Industrial Engineering, Universitat Politècnica de Catalunya

Dr. Fotini Labropulu
Mathematics - Luther College
University of ReginaPh.D., M.Sc. in Mathematics
B.A. (Honors) in Mathematics
University of Windso

Dr. Lynn Lim
Reader in Business and Marketing
Roehampton University, London
BCom, PGDip, MBA (Distinction), PhD, FHEA

Dr. Mihaly Mezei
ASSOCIATE PROFESSOR
Department of Structural and Chemical Biology, Mount Sinai School of Medical Center
Ph.D., Etvs Lornd University
Postdoctoral Training,
New York University

Dr. Söhnke M. Bartram
Department of Accounting and FinanceLancaster University Management SchoolPh.D. (WHU Koblenz)
MBA/BBA (University of Saarbrücken)

Dr. Miguel Angel Ariño
Professor of Decision Sciences
IESE Business School
Barcelona, Spain (Universidad de Navarra)
CEIBS (China Europe International Business School).
Beijing, Shanghai and Shenzhen
Ph.D. in Mathematics
University of Barcelona
BA in Mathematics (Licenciatura)
University of Barcelona

Philip G. Moscoso
Technology and Operations Management
IESE Business School, University of Navarra
Ph.D in Industrial Engineering and Management, ETH Zurich
M.Sc. in Chemical Engineering, ETH Zurich

Dr. Sanjay Dixit, M.D.
Director, EP Laboratories, Philadelphia VA Medical Center
Cardiovascular Medicine - Cardiac Arrhythmia
Univ of Penn School of Medicine

Dr. Han-Xiang Deng
MD., Ph.D
Associate Professor and Research
Department Division of Neuromuscular Medicine
Davee Department of Neurology and Clinical NeuroscienceNorthwestern University
Feinberg School of Medicine

Dr. Pina C. Sanelli
Associate Professor of Public Health
Weill Cornell Medical College
Associate Attending Radiologist
NewYork-Presbyterian Hospital
MRI, MRA, CT, and CTA
Neuroradiology and Diagnostic Radiology
M.D., State University of New York at Buffalo,School of Medicine and Biomedical Sciences

Dr. Roberto Sanchez
Associate Professor
Department of Structural and Chemical Biology
Mount Sinai School of Medicine
Ph.D., The Rockefeller University

Dr. Wen-Yih Sun
Professor of Earth and Atmospheric SciencesPurdue University Director
National Center for Typhoon and Flooding Research, Taiwan
University Chair Professor
Department of Atmospheric Sciences, National Central University, Chung-Li, TaiwanUniversity Chair Professor
Institute of Environmental Engineering, National Chiao Tung University, Hsin-chu, Taiwan.Ph.D., MS The University of Chicago, Geophysical Sciences
BS National Taiwan University, Atmospheric Sciences
Associate Professor of Radiology

Dr. Michael R. Rudnick
M.D., FACP
Associate Professor of Medicine
Chief, Renal Electrolyte and Hypertension Division (PMC)
Penn Medicine, University of Pennsylvania
Presbyterian Medical Center, Philadelphia
Nephrology and Internal Medicine
Certified by the American Board of Internal Medicine

Dr. Bassey Benjamin Esu
B.Sc. Marketing; MBA Marketing; Ph.D Marketing
Lecturer, Department of Marketing, University of Calabar
Tourism Consultant, Cross River State Tourism Development Department
Co-ordinator , Sustainable Tourism Initiative, Calabar, Nigeria

Dr. Aziz M. Barbar, Ph.D.
IEEE Senior Member
Chairperson, Department of Computer Science
AUST - American University of Science & Technology
Alfred Naccash Avenue – Ashrafieh

CONTENTS OF THE VOLUME

Global Journal of Computer Science and Technology
Volume 12 Issue 5 Version 1.0 March 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Fixed Mobile Convergence – Ims Approach

By Ashwini Patil & H K Sawant
Bharati Vidyapeeth Deemed University College Of Engineering, Pune

Abstract - The paper is aimed at studying and analyzing the network performance parameters of SIP protocol. SIP is content based protocol, in which various message are required to be transacted so that a session could be created, terminated or modified. Therefore, the objective is to analyze various SIP activities and the delay incurred in session start-up under various network conditions. Proper functioning of IMS platform is dependent on optimum performance of several protocols specified in the standard. Nearly all of the protocols used in IMS are standardized by the IETF. Some of the major protocols are SIP, SDP- signaling protocol, DIAMETER- improvised version of RADIUS protocol,COPS- Common Open Policy Service, H.248- descendant of MEGACo, RTP/RTCP- Real Time Protocol/Real Time Control Protocol, etc. Out of all these, Session Initiation Protocol (SIP) is the prominent protocol used to create, terminate and modify the sessions initiated by the user. In order to improve the performance parameters, this is the area where most of the research work is centralized. Hence, to study various aspects of SIP protocol with respect to the network performance is of great interest.

GJCST Classification: C.2.1

Strictly as per the compliance and regulations of:

Fixed Mobile Convergence – Ims Approach

Ashwini Patil [α] & H K Sawant [α]

Abstract - The paper is aimed at studying and analyzing the network performance parameters of SIP protocol. SIP is content based protocol, in which various message are required to be transacted so that a session could be created, terminated or modified. Therefore, the objective is to analyze various SIP activities and the delay incurred in session start-up under various network conditions. Proper functioning of IMS platform is dependent on optimum performance of several protocols specified in the standard. Nearly all of the protocols used in IMS are standardized by the IETF. Some of the major protocols are SIP, SDP- signaling protocol, DIAMETER- improvised version of RADIUS protocol,COPS- Common Open Policy Service, H.248- descendant of MEGACo, RTP/RTCP- Real Time Protocol/Real Time Control Protocol, etc. Out of all these, Session Initiation Protocol (SIP) is the prominent protocol used to create, terminate and modify the sessions initiated by the user. In order to improve the performance parameters, this is the area where most of the research work is centralized. Hence, to study various aspects of SIP protocol with respect to the network performance is of great interest.

I. Introduction

In today's market scenario, it is seen that the mobile segment is growing with the faster speed all over the globe. Considering the Indian telecom sector, it is noticed that our telecom network is the second largest network in the world and it is the fastest growing network in the world. That's why Indian telecom sector is the true representative of the whole world. Following important issues are cropping up in the telecom sector and constraining the telecom growth.

a) Volume of telecom traffic is rising day by day and tariffs are falling down abysmally low.
b) Launching of high data speed services like 3G and BWA is on the anvil. Content based services and VAS are on the rise.
c) Fixed line network is reducing and competing with the wireless services. The fixed line broadband growth is not able to meet the market demand. There is a scarcity of spectrum and it is adding more expenditure towards erection of extra infrastructure/BTS's in a defined geographical area. Call drops are more due to multiple hand-offs in small area, which is degrading the Quality of Service (QoS).

Broadband on WLAN's is an integral part of the Fixed wireless Network, which can offer connectivity at outdoor as well as indoor, has not established the roots prominently in many countries till today. Competitive pricing and aggressive marketing are greatly affecting the churning of customers. The solution to all these problems is the Convergence of Fixed and Wireless services.

Author [α] : Department of Information Technology, Bharati Vidyapeeth Deemed University College Of Engineering, Pune-46
E-mail : hemant1960@gmail.com

a) Present Trends

Excluding the demarcations evolving between business and personal lives, new styles are emerging, which are demanding for services that are required at business and personal level. Now a days short message can be sent between mobile devices and fixed line phones. Service providers can provide seamless services over the common network infrastructure on wireless and wire-line broadband networks. The implementation of IP and SIP makes it possible to establish multiple sessions over an IP network or multiple networks and gives perception that there is a single network.

b) Convergence

With the technological advancements, the differences between fixed and wireless networks and hence the respective services are becoming unnoticeable. This convergence can allow a mobile call of a user to be delivered on s fixed phone or a fixed phone call can be delivered on a mobile device as per the suitability of the user. The main objective being that the mobile services and fixed services can be established on a single device which can switch between networks. Fixed Mobile Convergence can be achieved via user mobile device which can support wide area (cellular) access and local area (Wi-Fi) access. The concept of convergence is more user centric than network access technology.

As per the ITU-T recommendation Q.1761 on the principle and requirements for convergence of fixed and existing IMT-2000 system defines Fixed Mobile Convergence (FMC) as "A mechanism by which, on IMT-2000 user can have his basic voice as well as other services through a fixed network as per his subscription options and capability of the access technology".

c) Convergence Characteristics

Converged Customer Premises Equipment (CPE) – e.g. The services that a user is accessing through different gadgets can be made available on a single Wi-Fi phone. Thus the Wi-Fi phone can become the converged CPE. Convergence brings up the concept of personalization of services in fixed line services as these services are used at par with the wireless services. Irrespective of the configuration, the converged services bring out the unification of different

aspects like freedom of mobility with the security, quality of service, higher bandwidth and lower costs of fixed-line services. It gives uniform communication experience at home and away from home with convenience and freedom of movement.

The converged network makes use of various radio spectrums and different technologies as a backbone infrastructure. The multi-radio infrastructure enables the cooperation of existing radio networks to combine their spectrum-efficient capabilities, whereby high-quality mobile multimedia services shall be provided.

II. Literature Survey

In the past few years, the evolution of cellular networks has reflected immense success and growth which was experienced by Internet in the last decade. This leads to networks where Internet Protocol connectivity is provided to mobile nodes. The result is third generation (3G) networks where IP services such as voice over IP (VoIP) and instant messaging (IM) are provided to mobile nodes (MN) in addition to connectivity. With import of Wi-Fi, Wi-MAX, digital video broadcasting, satellite, internet, etc. the current architecture of telecommunication network has been facing a challenge of interoperability.

Hence, the solution was to devise an interoperable platform which can provide the services irrespective of the access technology. The basic approaches to convergence are

UMA: Unlicensed Mobile Access (UMA) is a new technology that provides access to GSM services over Wireless LAN or Bluetooth. It also challenges the assumption of closed platform, since it is relatively easy to implement a UMA phone purely in software running on standard PC hardware and operating systems. In the UMA solution, exiting cellular network remains unmodified, and a new network element, the UMA Network Controller (UNC), is introduced. UNC acts as a gateway between the mobile operator core network and Internet or a broadband IP access network such as ADSL or cable. The phone connects to the IP network using a standard WLAN or Bluetooth access point. Since GSM/GPRS core security mechanisms; new mechanisms are defined only for protecting the communication between the phone and UNC.

SIP : The Session Initiation Protocol (SIP) is a signaling protocol used for establishing sessions in an IP network. A session could be a simple two-way telephone call or it could be a collaborative multi-media conference session. The ability to establish these sessions means that a host of innovative services become possible, such as voice-enriched e-commerce, web page click-to-dial, Instant Messaging with buddy lists, and IP Centrex services. SIP is a request-response protocol that closely resembles two other internet protocols, HTTP and SMTP; consequently, SIP sits comfortably alongside Internet application. Using SIP, telephony becomes another web application and integrates easily into other Internet services. SIP is a simple toolkit that service providers can use to build converged voice and multimedia services. SIP is an "application-layer control (signaling) protocol for creating, modifying, and terminating sessions with one or more participants. These sessions include Internet telephone calls, multimedia distribution, and multimedia conferences."

IMS: IP Multimedia Subsystem (IMS) is referred as the heart of NGN. The 3GPP has published a number of specifications that define the IMS as the part of the 3G wireless environment which will "enable the convergence of, and access to, voice, video, messaging, data and web-based technologies for the wireless user." The 3GPP defined the IMS specifications in support of the Universal Mobile Telecommunication System (UMTS). The UMTS is the 3G evolution of the GSM.

IMS is the envisioned solution that will provide new multimedia rich communication services by mixing telecom and data on an access independent IP based architecture, defined in 3rd Generation Partnership Project (3GPP), 3rd Generation Partnership Project 2 (3GPP2) and Internet Engineering Task Force (IETF) standards.

The aim of IMS is to provide all the services, current and future, that the Internet provides with roaming facilities. To achieve these goals, IMS supports peer-to-peer IP communications between existing technology standards while providing a framework for inter-operability of voice and data services for both fixed (POTS, ISDN) and mobile users (802.11, GSM, CDMA, UMTS). It provides session control, connection control and an application services framework with both subscriber and services data, while allowing interoperability of these converged services between subscribers. IMS truly merges the Internet with the cellular world; it uses cellular technologies to provide ubiquitous access and Internet technologies to provide appealing services.

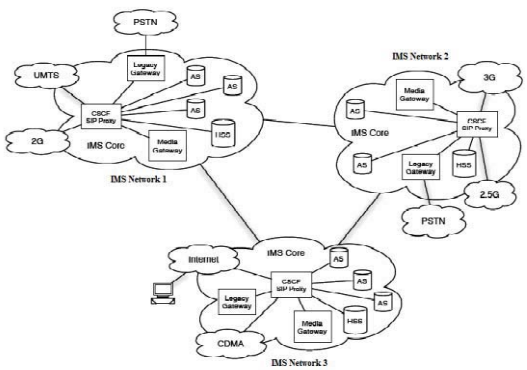


Figure:1 IMS Overview : NGN Convergence

III. Layered Architecture of Ims

The definition of IMS by 3GPP is an all packet core network that is able to accept all types of access networks i.e. WiMAX, CDMA, GSM in order to deliver a wide range a multimedia services which can be offered to the user by any device connected to the network. The use of SIP in IMS enable the support of IP-to-IP sessions over any wire-line connection system like DSL as well as wireless networks like Wi-Fi, GSM and CDMA. The IMS allows the inter-working between the traditional TDM networks and the IP networks.

a) Access Layer

The IMS architecture is independent of any access bearer. In the mobile networks the access layer could be any or a combination of the following: General Packet Radio Service (GPRS), Enhance Data Rate for GSM Evolution (EDGE), Code Division Multiple Access (CDMA), Wireless Interoperability for Microwave Access (WiMAX), Universal Mobile Telecommunication System (UMTS) and Wireless Local Area Networks (W-LAN or Wi-Fi). The fixed line networks makes use of Asymmetric Digital Subscriber Lines (ADSL) and cable network accesses.

b) Transport Layer

This is an all IP network which consist of IP Routers. These routers are Label Edge Routers and Core Switching Networks. The IP/MPLS (Multi-protocol cable Switching) is the transport layer technology for IMS platform. MPLS defines a mechanism for the forwarding of packets in a router network. Due to its flexibility, the MPLS has become the default IP transport network for the Next Generation Networks making use of the IMS core in order to reliability and Quality of Service.

c) Session Control Layer

This layer consists of network control servers which are used for the management of calls, establishment and modifications of sessions in the IMS platform. Two main elements in this layer are the Call Session Control Function (CSCF) and the Home Subscriber Server (HSS). These two elements form the core of the ISM architecture and are sometimes referred to as SIP Servers. The CSCF provide end-point registration and routing of SIP signaling messages and provide interworking with the transport layer for guaranteed QoS of all services. The HSS is the database that is used to store the subscriber service profiles and service triggers. The other information stored by the HSS includes the dynamic data of the subscribes like location information.

d) Application Layer

This layer makes use of application and content servers to provide value-added services. The Application Server (AS), the Multimedia Resource Function Controller (MRFC) and Multimedia Function Resource processor (MFRP) form the core of this layer. The AS is responsible for the execution of service-specific logic i.e. user interaction with subscribers and call flows. The MRFP is also known as IP media server is used to provide media processing for the application layer. The media server is used to enable to delivery of some non-telephony services like Push-to-Talk (PTT), speech enabled services, video services and other services like conferencing, prepaid and personalized call-back tones.

The control and application layers are access and transport independent and this is very useful in order to ensure that the user is able access the ISM services required from any access network and from any connection device.

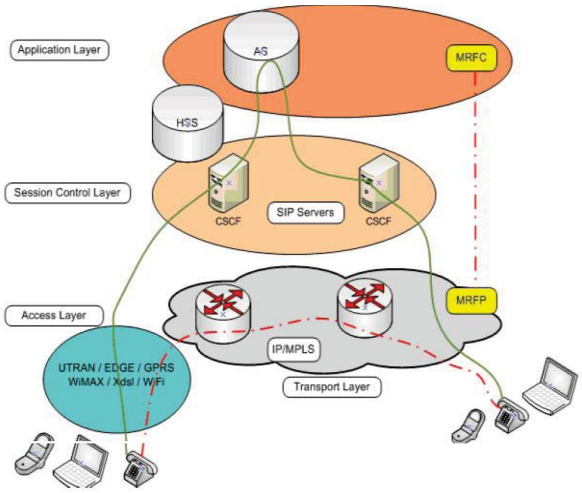


Figure 2 Layered Architecture of IMS

IV. Ims Architecture Design

In the General architecture of IMS, 3GPP standardize functions but not nodes. So we can say that the IMS architecture is a collection of functions linked by standardized interfaces. If any implementer want they can merge two functions into single node as well as they can divide single function into two or more nodes. An overview of the IMS architecture as standardized by 3GPP.[1,2] Here we include only most important nodes. The common nodes included in the IMS are as follows:

a) CSCF (Call/Session Control Function)

CSCF is a SIP (Session Initiation Protocol) server which processes SIP signaling in the IMS. CSCFs are dynamically associated, service-independent and standardized access points. It distributes incoming calls to the application services and handles initial subscriber authentication. There are three types of CSCFs depending on the functionality they provide.

P-CSCF (Proxy-CSCF): The P-CSCF is the first point of contact between the IMS terminal and the IMS network. All the requests initiated by the IMS terminal or

destined to the IMS terminal traverse the P-CSCF. This node provides several functions related to security. The P-CSCF also generates charging information toward a charging collection node. An IMS usually includes a number of P-CSCFs for the sake of scalability and redundancy. Each P-CSCF serves a number of IMS terminals, depending on the capacity of the node.

I-CSCF (Interrogating-CSCF): The I-CSCF provides the functionality of a SIP proxy server. It also has an interface to the SLF (Subscriber Location Function) and HSS (Home Subscriber Server). This interface is based on the Diameter protocol (RFC 3588). I-CSCF retrieves user location information and routes the SIP request to the appropriate destination, typically an S-CSCF.

S-CSCF (Serving-CSCF): The S-CSCF is a SIP server that performs session control. It maintains a binding between the user location and the user's SIP address of record (also known as Public User Identity). Like the I-CSCF, the S-CSCF also implements a Diameter interface to the HSS.

b) SIP AS (Application Server)

The AS is a SIP entity that hosts and executes IP Multimedia Services based on SIP.

c) MGCF (Media Gateway Control Function)

MGCF implements a state machine that does protocol conversion and maps SIP to either ISUP (ISDN User part) over IP or BICC (Bearer Independent Call Control) over IP. The protocol used between the MGCF and the MGW is H.248 (ITU-T Recommendation H.248).

d) MGW (Media Gateway)

The MGW interfaces the media plane of the PSTN (Public Switched Telephone Network) or CS (Circuit Switched) network. On one side the MGW is able to send and receive IMS media over the Real-Time Protocol (RTP). On the other side the MGW uses one or more PCM (Pulse Code Modulation) time slots to connect to the CS network. Additionally, the MGW performs trans-coding when the IMS terminal does not support the codec used by the CS side.

e) HSS (Home Subscriber Server)

It contains all the user related subscription data required to handle multimedia sessions. These data include, among other items, location information, security information (including both authentication and authorization information), user profile information and the S-CSCF allocated to the user. The SLF (Subscription Location Function) is a simple database that maps users' addresses to HSSs. Both the HSS and the SLF implement the Diameter protocol.

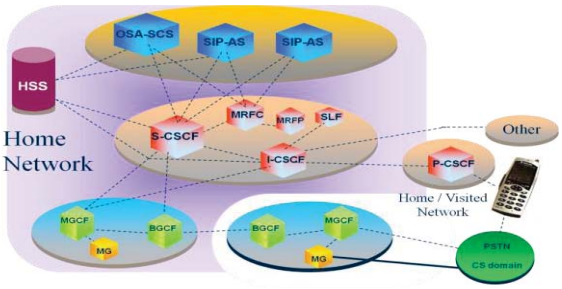


Figure 3 IMS Architecture

V. System Description

Several works analyzed the subject of SIP signaling from different angles. All these works did not associate any constraints to the access network and assumed an infinite bandwidth over the link between the User Agents, IMS entities and intermediate routers. This assumption is not true and the impact of SIP signaling on the network and its consequences on the SIP nodes must be evaluated. Apart from this, most of the studies were carried out with topology containing either fixed nodes or mobile nodes only; therefore the performance of a hybrid topology can be studied too.

So keeping in mind the drawback of the previous works, we propose a topology consisting of both mobile and fixed nodes with proper specification of suitable links between them.

Statistic to Be Obtained

1. Tunneled Traffic Sent/Received: Amount of traffic tunneled by Agents and de-tunneled by agents and hosts. It is given in packets/sec.
2. Jitter: If two consecutive packets leave the source node with time stamps t1 & t2 and are played back at the destination node at time t3 & t4, then: Jitter = (t4 - t3) - (t2 - t1) Negative jitter indicates that the time difference between the packets at the destination node was less than that at the source node. It is given in sec.
3. Packet End to End Delay: The total voice packet delay, called "mouth-to-ear" delay = network delay + encoding delay + decoding delay + compression delay + decompression delay. Specified in seconds.
4. Wireless LAN Delay: Represents the end to end delay of all the packets received by the wireless LAN MACs of all WLAN nodes in the network and forwarded to the higher layer. Specified in seconds.
5. Retransmission Attempts: Total number of retransmission attempts by all WLAN MACs in the network until either packet is successfully transmitted or it is discarded as a result of reaching short or long retry limit. Specified in packets.
6. Registration Traffic Sent/Received: Amount of Mobile IP registration packets sent/received. Mobile IP nodes and router register their addresses with

agent nodes (home/foreign) to receive mobile ip service.

7. Throughput: This statistic represents the average number of packets successfully received or transmitted by the receiver or transmitter channel per second.
8. Utilization: This statistic represents the percentage of occupancy of an available channel bandwidth with respect to time period, where a value of 100.0 would indicate full usage.
9. SIP Active Calls: Number of active calls at any given time.
10. Call Setup Time: Time to setup a call in seconds.
11. Calls Connected: It includes calls initiated by this node and also the incoming call requests.
12. Calls Initiated: Number of calls initiated at a particular SIP UAC node.
13. Call Duration: Duration of each call defined as the time at which the node got call connect confirmation to the time at which it got call disconnect confirmation.

VI. Conclusion

Next Generation Networks (NGNs) aims at providing a wide range of services to end-users over an access independent platform while allowing for better Quality of service (QoS), charging mechanism and integration of services as compared to conventional fixed or mobile networks. The NGN core network is known as IP Multimedia Subsystem (IMS) which is the Third Generation Partnership Project (3GPP) standardized core network for all IP-convergence of fixed and mobile networks.

The core functionality of the IMS is built on the Session Initiation Protocol (SIP), the Internet Engineering Task Force (IETF) standardized protocol for the creation, management and termination of multimedia sessions on the internet. Hence to study the signalling of SIP traffic is the subject of major interest.

In this work we have to analyze the impact of different types of traffic load with varying pattern of the SIP signalling carried over the network. We need to configure a hybrid network topology consisting of IMS entities I-CSCF, P-CSCF, S-CSCF, intermediate routers and SIP enabled fixed and mobile nodes. Then we have to create different scenarios and varied their respective parameters. Other signalling protocol like H.323, etc, do not support mobility but SIP handles this problem because of the Mobile IP support.

The IMS is based on Session Initiation Protocol (SIP) which is a text based protocol. The IMS will generally create additional signaling traffic in the IP based networks, so there is a need to take necessary precautions to minimize the signaling overload.

References Références Referencias

1. 3GPP, TSG SSA, IP Multimedia Subsystem (IMS) – Stage 2 (Release 7), TS23.228 v.7.3.0, 2006-03.
2. 3GPP, Technical Specification Group Services and System Aspects, IP
3. Multimedia Subsystem (IMS) - Stage 2 (Release 5), TS 23.228 v5.6.0, 2002-09.
4. 3GPP, Technical Specification Group Services and System Aspects: QoS Concept and
5. Architecture (Release 5), TS 23.207 v5.6.0, 2002-09.
6. 3GPP, Technical Specification Group Core Network; Signalling flows for the IP multimedia call control based on SIP and SDP; Stage 3 (Release 5), TS 24.228 v5.2.0, 2002-09.
7. Perkins, "IP mobility support for IPv4," RFC 3344, Internet Engineering Task Force, 2002.
8. V. Planat, N. Kara; “SIP Signaling retransmission analysis over 3G network”, Proceedings of MoMM 2006
9. Jie Xiao, Changcheng Huang and James Yan; “A Flow-based Traffic Model for SIP Messages in IMS”, 978-1-4244-4148-8/09 IEEE "GLOBECOM" 2009 proceedings
10. Arslan Munir, Gordon Ross, “SIP-Based IMS Signaling Analysis for WiMax-3G Interworking Architectures”, 1536-1233/10 IEEE Transactions on mobile computing, Vol. 9, No. 5, May 2010.
11. 3GPP. TS 24 229”IP Multimedia Call Control Protocol based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP)” V5.16.0, March 2006.
12. Session Initiation Protocol (SIP), RFC 3261, H. Schulzrinne, J. Rosenberg. June 2002
13. GPP. TS 24 228 Signalling flows for the IP multimedia call control based on Session Initiation Protocol (SIP) and Session Description Protocol (SDP) V5.14, December 2005.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 5 Version 1.0 March 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Security Enhancement of E-Voting System

By Sanjay Kumar & Manpreet Singh
M. M. University, Ambala

Abstract - The term E-Voting is used in variety of different ways and it encompasses all voting techniques involving electronic voting equipments, voting over the internet, using electronic booths in polling stations and sometimes even counting of paper ballots. A voting system that can be proven correct has many concerns. The basic reasons for a government to use electronic systems are to increase election activities and to reduce the election expenses. Still there is some scope of work in electronic voting system in terms of checking the authenticity of voters and securing electronic voting machine from miscreants. Biometrics is automated tool for verifying the identity of a person based on a physiological or behavioral characteristic. It has the capability to reliably distinguish between an authorized person and an imposter. Since biometric characteristics are distinctive, can not be forgotten or lost and the person to be authenticated needs to be physically present at the point of identification, biometrics is inherently more reliable and more capable than traditional knowledge-based and token-based techniques. In this paper, we have proposed a model to enhance the security of electronic voting system by incorporating fast and accurate biometric technique to prevent an unauthorized person to vote.

Keywords : *E-Voting, Biometric, Fingerprint Recognition.*

GJCST Classification: *K.4.0*

Strictly as per the compliance and regulations of:

Security Enhancement of E-Voting System

Sanjay Kumar [α] & Manpreet Singh [σ]

Abstract - The term "E-Voting" is used in variety of different ways and it encompasses all voting techniques involving electronic voting equipments, voting over the internet, using electronic booths in polling stations and sometimes even counting of paper ballots. A voting system that can be proven correct has many concerns. The basic reasons for a government to use electronic systems are to increase election activities and to reduce the election expenses. Still there is some scope of work in electronic voting system in terms of checking the authenticity of voters and securing electronic voting machine from miscreants. Biometrics is automated tool for verifying the identity of a person based on a physiological or behavioral characteristic. It has the capability to reliably distinguish between an authorized person and an imposter. Since biometric characteristics are distinctive, can not be forgotten or lost and the person to be authenticated needs to be physically present at the point of identification, biometrics is inherently more reliable and more capable than traditional knowledge-based and token-based techniques. In this paper, we have proposed a model to enhance the security of electronic voting system by incorporating fast and accurate biometric technique to prevent an unauthorized person to vote.

Keywords : *E-Voting, Biometric, Fingerprint Recognition.*

I. Introduction

Electronic Voting Machine (EVM) is a simple electronic device used to record votes in place of ballot papers and boxes which were used earlier in conventional voting system. It is a simple machine that can be operated easily by both the polling personnel and the voters. Being a standalone machine without any network connectivity, nobody can interfere with its programming and manipulate the results. Advantages of EVM [1] over the traditional ballot paper/ box system are:

- It eliminates the possibility of invalid and doubtful votes which, in many cases, are the root causes of controversies and election petitions.
- It makes the process of counting of votes much faster than the conventional system.
- It reduces to a great extent the quantity of paper used thus saving a large number of trees making the process eco-friendly.
- It reduces cost of printing almost nil as only one sheet of ballot paper is required for each polling.

II. Present Voting System

Voting is the bridge between the governed and government. The last few years have brought a renewed focus onto the technology used in the voting process and a hunt for voting machines. Computerized voting systems bring improved usability and cost benefits but suffer from weak software which has lot of bugs. When scrutinized, current voting systems have security holes and it becomes difficult to prove even simple security properties about them. A voting system that can be proven correct would solve many problems.

High security is essential to elections. There has been a lot of attention to an electronic voting by cryptographers. Many scientific researches have been done in order to achieve security, privacy and correctness in electronic voting systems by improving cryptographic protocols of e-voting systems. Currently, the practical security in e-voting systems is more important than the use of cryptographic schemes [2]. One of the main interests is seemingly contradicting security properties. On one hand, voting must be private and the votes should be anonymous. On the other hand, voters must be identified in order to guarantee that only the eligible voters are capable to vote. Hence, e-voting should be uniform, confidential, secure and verifiable. The most important requirements for e-voting can be characterized as:

- Eligible voter is authenticated by his/her unique characteristics.
- Eligible voters are not allowed to cast more than one vote.
- Votes are secret.
- Auditors can check whether all correct cast ballots participated in the computation of the final tally.
- Result of election should be secret until the end of an election.
- While voting is on, there should not be a method of knowing intermediate result that can affect the remaining voter's decisions.
- All valid votes must be counted correctly and the system outputs the final tally.
- It must be possible to repeat the computation of the final tally.

The following three dimensions are used to make a comparison of electronic voting systems for various nations [6]:

- Whether a country's system uses a paper audit trail.
- Whether the system permits an anonymous, blank or spoiled ballot.
- Whether the software is open source or proprietary.

III. Proposed E-Voting System

We have proposed a model for e-voting based on biometric technique. Biometrics has been widely used in various applications such as criminal

identification, prison security electronic banking, e-commerce [3]. Biometric authentication requires comparing a registered or enrolled biometric sample (biometric template or identifier) against a newly captured biometric sample (for example, a fingerprint captured during a login). During enrollment, a sample of the biometric trait is captured, processed by a computer and stored for later comparison. Biometric recognition can be used in identification mode, where the biometric system identifies a person from the entire enrolled population by searching a database for a match based solely on the biometric. A system can also be used in verification mode, where the biometric system authenticates a person's claimed identity from their previously enrolled pattern. This is also called "one-to-one" matching [4]. The proposed model uses biometrics in the verification mode during e-voting. Implication of error rates of different biometric techniques based on False Reject Rate (FRR) and False Acceptance Rate (FAR), we can conclude that the finger print recognition is that fast and accurate biometric technique required for making reliable and secure system [7].

The patterns of friction ridges and valleys on an individual's fingertips are unique to that individual. For decades, law enforcement has been classifying and determining identity by matching key points of ridge endings and bifurcations. Fingerprints are unique for each finger of a person including identical twins. One of the most commercially available biometric technologies, fingerprint recognition devices for desktop and laptop access are now widely available, users no longer need to type passwords – instead, only a touch provides instant access. Fingerprint systems can also be used in identification mode. Several states check fingerprints for new applicants to social services benefits in order to ensure that recipients do not fraudulently obtain benefits under fake names [5]. Fingerprints are the ridge and furrow patterns shown in Figure 1 on the tip of the finger and have been used extensively for personal identification of people. The availability of cheap and compact solid state scanners as well as robust fingerprint matchers are two important factors in the popularity of fingerprint-based identification systems

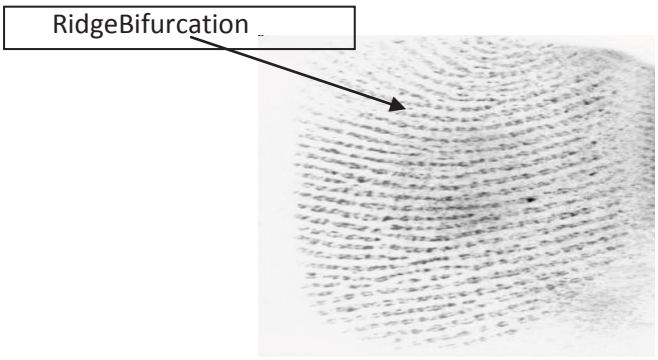


Fig.1. Ridges and Furrows

The proposed model assumes that the unique id has been assigned to each voter (Citizen of India). Each perspective voter must have a unique fingerprint registered in the database. The proposed model comprises of following steps:

Step-1: Open login window.

Step-2 : Enter P.O.id and password. /* id and password filled by Presiding Officer(P.O.) */

Step-3 : If P.O. id and password verifies
Main window /* main window will appear automatically */
Else
Print- wrong id and password
Go to step 2

Step-4: Enter user id and password. /* for voter, if the system has been successfully started by P.O. */
count=0;

Step-5 : Verify userid and password
If user id and password matches
If votedflag=false
Print- enter thumb impression
count++
If thumb impression valid
go to step 6
Else
Print-This time your thumb impression didn't match
Go to Block 1
Else
Print- you have already voted
Go to step 7
Else
Print Userid and password don't match
Go to step 7
Block 1 Print (chances left, "3-count")
If(count <=3)
Go to step 5
Print- You are not authorized
Go to step 7

Step-6 : Print- please enter your vote
Set votedflag = true
Print-thank you

Step-7 : Check current time (It should be less than closing time 05:00PM.)
If (current time < 05:00 PM)
Go to step 4 /* another user can vote*/
Else
Go to step 8
/*After 5:00 PM, no person is eligible to restart the system for voting procedure.*/

Step-8 : Exit

IV. Implementation

This proposed model has been simulated successfully on Java platform. Figure 2 to Figure 5 are

few snapshots of steps involved in the implementation of proposed model.

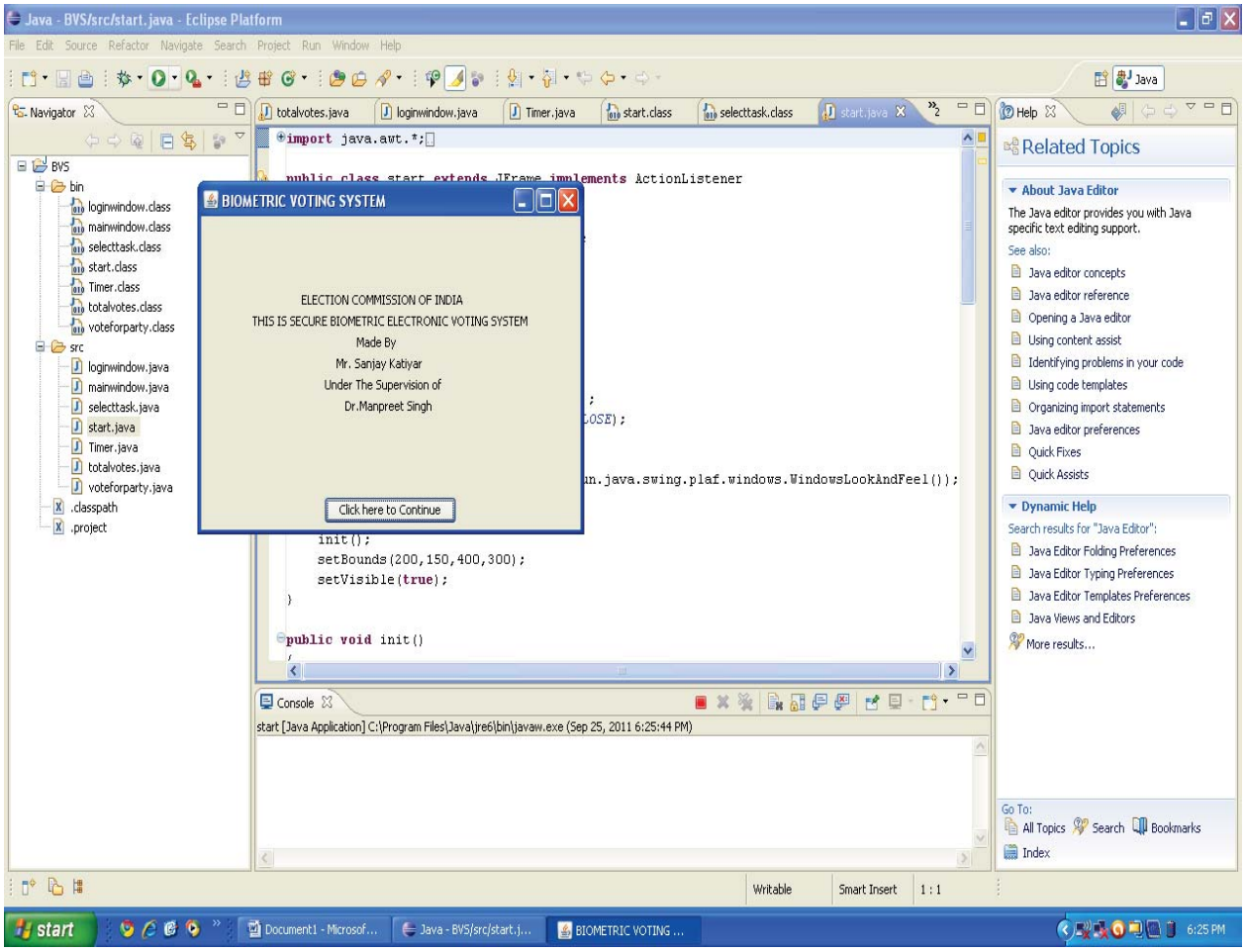


Fig.2. Home Page of Biometric Electronic Voting System

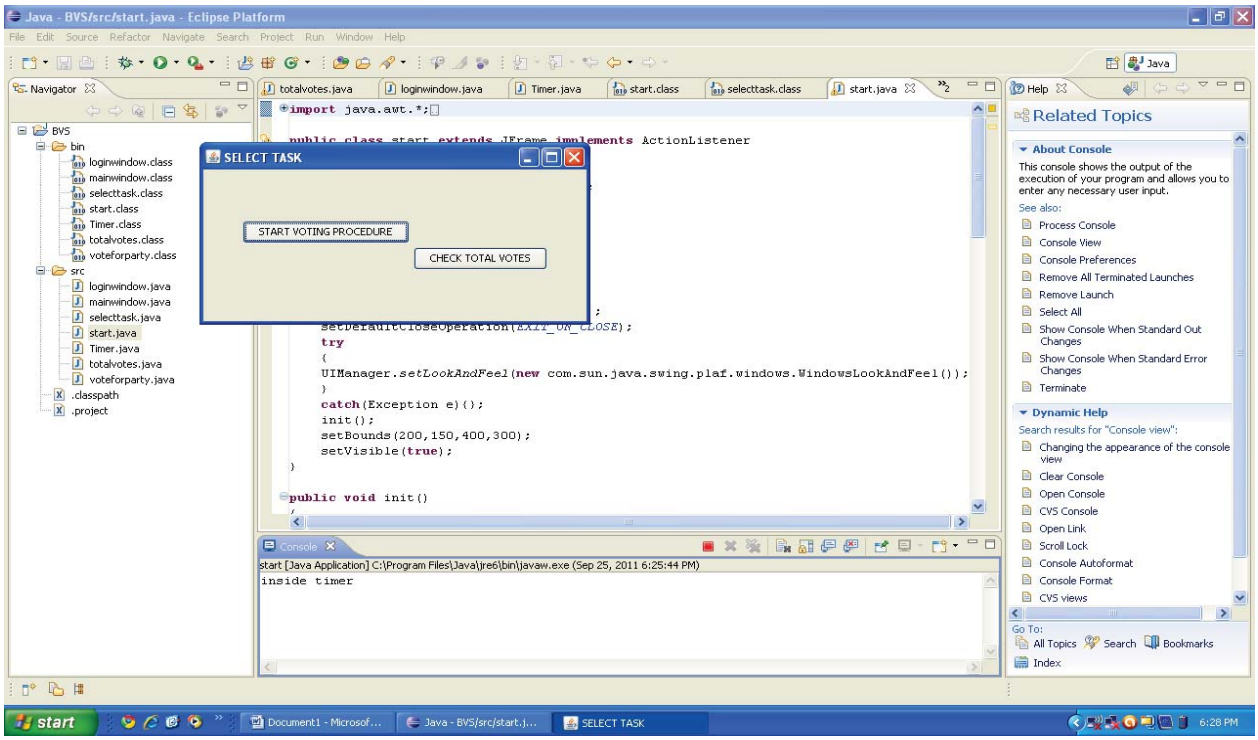


Fig.3 Login Window of Presiding Officer

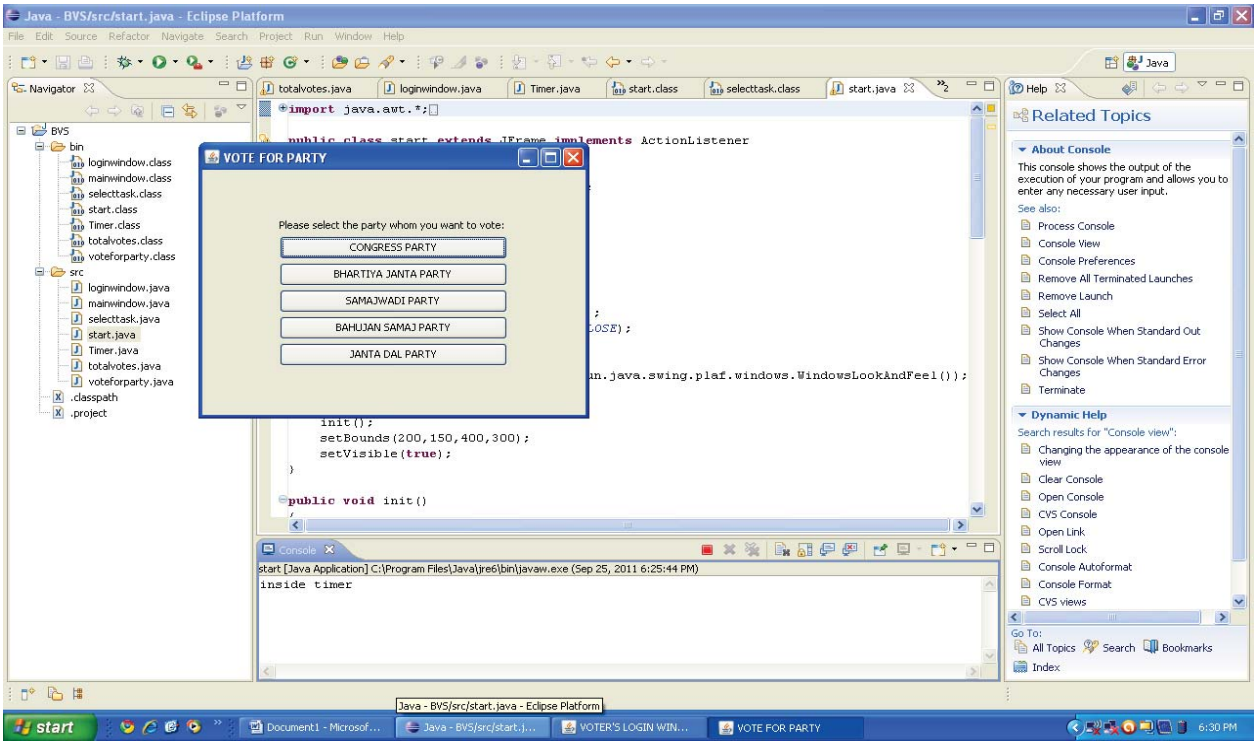


Fig.4. Voting Procedure

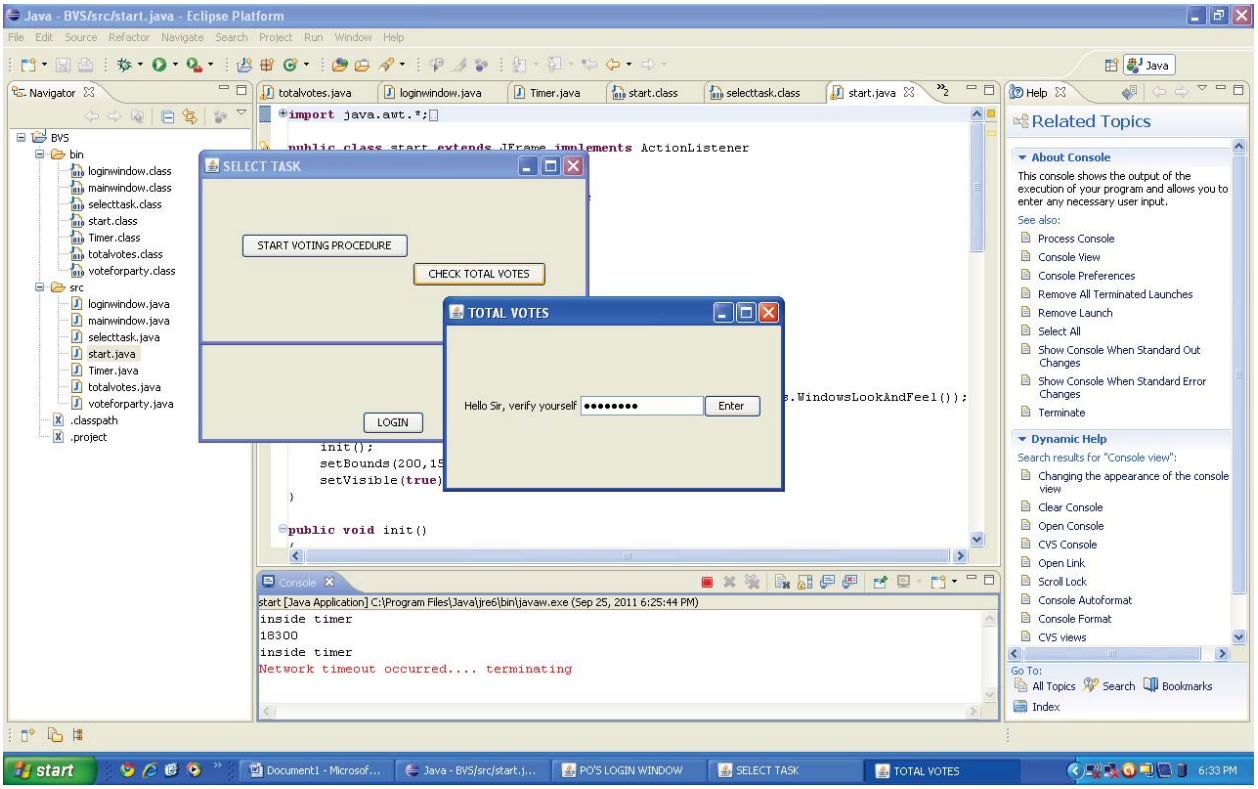


Fig.5. Vote Counting

We have implemented the proposed model by integrating it with FTA 5454(A10) - Fingerprint Time and Attendance system as shown in Figure 6. It can store 3000 fingerprint templates and 50000 transaction records.

Fig.6. Device used for Fingerprint Recognition

V. Conclusion

We have presented a model for electronic voting wherein fingerprint is embedded as biometrics for voter identification. The future work will concentrate on implementation of fast and accurate fingerprint recognition and other related technical aspects in the system.

References Références Referencias

1. Everett, S., Greene, K., Byrne, M., Wallach, D., Derr, K., Sandler, D., Torous, T. (2008). Electronic voting machines versus traditional methods: improved preference, similar performance. Proceeding of 26th Annual SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, April 5-10, 2008, 883-892.
2. Ansper, A., Buldas, A., Oruaas, M., Piirsalu, J., Veldre, A., Willemson, J., Kivinurm, K. (2002). The Security of Conception of E-Voting: Analysis and Measures. Technical Report by National Electoral Committee.
3. Phillips, P., Martin, A., Wilson, C., Przybocki, M. (2000). An Introduction to Evaluating Biometric Systems. IEEE Computer, 33(2), 56-63.
4. Zebbiche, K., Khelifi, F., Bouridane, A. (2008). An Efficient Watermarking Technique for the Protection of Fingerprint Images. EURASIP Journal on Information Security, 2(1), 1-20.
5. Carrigan, Robert, Milton, Ron, Morrow, Dan (2005). "Automated fingerprint identification systems", Computer World Honors Case Study, 1-5.
6. Kumar, Sanjay, Walia, Ekta (2011). Analysis of Electronic Voting system in Various Countries. International Journal on Computer Science and Engineering, 3(5), 1825-1830.
7. Kumar, Sanjay, Walia, Ekta (2011). Analysis of various Biometric Techniques. International Journal of Computer Science and Information Technologies, 2(4), 1595-1597.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 5 Version 1.0 March 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Vision-Based Deep Web Data Extraction for Web Document Clustering

By M. Lavanya & Dr.M.Usha Rani
Sree Vidyanikethan Engineering College

Abstract - The design of web information extraction systems becomes more complex and time-consuming. Detection of data region is a significant problem for information extraction from the web page. In this paper, an approach to vision-based deep web data extraction is proposed for web document clustering. The proposed approach comprises of two phases: 1) Vision-based web data extraction, and 2) web document clustering. In phase 1, the web page information is segmented into various chunks. From which, surplus noise and duplicate chunks are removed using three parameters, such as hyperlink percentage, noise score and cosine similarity. Finally, the extracted keywords are subjected to web document clustering using Fuzzy c-means clustering (FCM).

Keywords : *Noise Chunk, cosine similarity, Title word Relevancy, Keyword frequency-based chunk selection, Fuzzy c-means clustering (FCM)*

GJCST Classification: *I.4.m, H.2.8*

Strictly as per the compliance and regulations of:

Vision-Based Deep Web Data Extraction for Web Document Clustering

M. Lavanya [α] & Dr.M.Usha Rani [σ]

Abstract - The design of web information extraction systems becomes more complex and time-consuming. Detection of data region is a significant problem for information extraction from the web page. In this paper, an approach to vision-based deep web data extraction is proposed for web document clustering. The proposed approach comprises of two phases: 1) Vision-based web data extraction, and 2) web document clustering. In phase 1, the web page information is segmented into various chunks. From which, surplus noise and duplicate chunks are removed using three parameters, such as hyperlink percentage, noise score and cosine similarity. Finally, the extracted keywords are subjected to web document clustering using Fuzzy c-means clustering (FCM).

Keywords : *Noise Chunk, cosine similarity, Title word Relevancy, Keyword frequency-based chunk selection, Fuzzy c-means clustering (FCM)*

I. Introduction

Today, World Wide Web has become one of the most significant information resources. Though most of the information is in the form of unstructured text, a huge amount of semi-structured objects, called data records, are enclosed on the Web [5]. Due to the heterogeneity and lack of structure of Web information, automated discovery of relevant information becomes a difficult task [1]. The Deep Web is the content on the web not accessible by a search on general search engines, which is also called as hidden Web or invisible Web[4]. Deep Web contents are accessed by queries submitted to Web databases and the retrieved information i.e., query results is enclosed in Web pages in the form of data records. These special Web pages are generated dynamically and are difficult to index by conventional crawler based search engines, namely Google and Yahoo. In this paper, we describe this kind of special Web pages as deep Web pages [12]. In general, Web information extraction tools are divided into three categories: (i) Web directories, (ii) Meta search engines, and (iii) Search engines.

In addition to main content, web pages usually have image-maps, logos, advertisements, search boxes, headers and footers, navigational links, related links and copyright information in conjunction with the main content. Though these items are required by web site owners, they will obstruct the web data mining and decrease the performance of the search engines [14], [15]. Hence, having a method that automatically discovers the information in a web page and allots substantial measures for different areas in the web page is of an immense advantage [19], [20]. It is imperative to distinguish relevant information from noisy content because the noisy content may deceive users' concentration within a solitary web page, and users only pay attention to the commercials or copyright when they search a web page.

Clustering is a technique, in which the data objects are given into a set of disjoint groups called clusters so that objects in each cluster are more analogous to each other than the objects from different clusters. Clustering techniques are used in several application areas such as pattern recognition (Webb, 2002), data mining (Tan, Steinbach, & Kumar, 2005), machine learning (Alpaydin, 2004), and so on. Generally, clustering algorithms can be classified as Hard, Fuzzy, Possibilistic, and Probabilistic[2] (Hathway & Bezdek, 1995).

In this paper a novel method to extract data items from the deep web pages automatically is proposed. It comprises of two steps: (1) Identification and Extraction of the data extraction for deep web page (2) Web clustering using FCM algorithm. Firstly in a web page, the irrelevant data such as advertisements, images, audio, etc are removed using chunk segmentation operation. The result we will obtain is a set of chunks[3]. From which, the surplus noise and the duplicate chunks are removed by computing the three parameters, such as *Hyperlink percentage, Noise score* and *cosine similarity*. For each chunk, three parameters such as *Title word Relevancy*, Keyword *frequency based chunk selection* and *Position feature* are computed. These sub-chunks consider as the main chunk and the keywords are extracted from those main chunk. Secondly, the set of keywords are clustered using Fuzzy c-means clustering.

The paper is organized as follows. Section 2 presents the related works. The problem statement is described in section 3 and the contribution of this paper is given in section 4. The definition of terms used in the proposed approach given in section 5. An efficient approach web document clustering based on vision-based deep web is discussed in section 6. The

Author α : Sr. Lecturer, SVEC, Tirupati Andhrapradesh, INDIA-51
E-mail : lavanya4_79@rediffmail.com
Author σ : Assoc .Prof, Dept. of C S, SPMVV, Tirupati , Andhrapradesh, INDIA- 517102
E-mail : musha_rohan@yahoo.com

experimental results are reported in Section 7. Section 8 explains conclusion of the paper.

II. Review of Related Works

Our proposed method concentrates on web document clustering based on vision-based deep web data extraction. Many Researchers have developed several approaches for web document clustering based on vision-based deep web data[7]. Among them, a handful of significant researches that performs web clustering and data extraction are presented in this section.

Moreover, a multi-objective genetic algorithm-based clustering method has been used for finding the number of clusters and the most natural clustering. It is complex and even impossible to employ a manual approach to mine the data records from web pages in deep web. Thus, *Chen Hong-ping et al* [9] have proposed a LBDRF algorithm to solve the problem of automatic data records extraction from Web pages in deep Web. Experimental result has shown that the proposed technique has performed well.

Zhang Pei-ying and Li Cun-he [10] have proposed a text summarization approach based on sentences clustering and extraction. The proposed approach includes three steps: (i) the sentences in the document have been clustered based on the semantic distance, (ii) the accumulative sentence similarity on each cluster has been calculated based on the multi-features combination technique, and (iii) the topic sentences has been selected via some extraction rules. The goal of their research is to exhibit that the summarization result was not only depends on the sentence features, but also depends on the sentence similarity measure. Qingshui Li and Kai Wu [6] have developed a Web Page Information extraction algorithm based on vision character. A vision character rule of web page has been employed, regarding the detailed problem of coarse-grained web page segmentation and the restructure problem of the smallest web page segmentation[8]. Then, the vision character of page block has been analyzed and finally determined the topic data region accurately.

ECON can be applied to Web news pages written in several well known languages namely Chinese, English, French, German, Italian, Japanese, Portuguese, Russian, Spanish, and Arabic. Also, ECON can been implemented without any difficulty. *Wei Liu et al [12]* have introduced a vision-based approach that is Web-page programming- language-independent for deep web data extraction. Mainly, the proposed approach has used the visual features on the deep Web pages to implement deep Web data extraction, such as data record extraction and data item extraction[11]. They have also proposed an evaluation measure revision to gather the amount of human effort required to produce proper extraction.

III. Problem Statement

In a web page, there are numerous immaterial components related with the descriptions of data objects. These items comprise advertisement bar, product category, search panel, navigator bar, and copyright statement, etc. Generally, a web page W_P is specified by a triple $W_P = (\omega, \phi, \eta)$. $\varpi = \{W_{p1}, W_{p2} \ldots W_{pn}\}$ is a finite set of objects or sub-web pages. All these objects are not overlapped. Each web page can be recursively viewed as a sub-web-page and has a subsidiary content structure. $\phi = \{\phi_1, \phi_2 \ldots \phi_n\}$ is a finite set of visual separators, such as horizontal separators and vertical separators. Every separator has a weight representing its visibility, and all the separators in the same ϕ have same weight. η is the relationship of every two blocks in ϖ, which is represented as: $\eta = \varpi \times \varpi \to \phi \cup \{NULL\}$. In several web pages, there are normally more than one data object entwined together in a data region, which makes it complex to find the attributes for each page. Also, since the raw source of the web page for representing the objects is non-contiguous one, the problem becomes more complicated. In real applications, the users necessitate from complex web pages is the description of individual data object derived from the partitioning of data region.

VI. Contribution of The Paper

We present new approach for deep web clustering based capture the actual data of the deep web pages. We achieve this in the following two phases. (1) Vision based Data relevant identification (2) Deep web pages clustering.

In the first phase,

- A data extraction based measure is also introduced to evaluate the importance of each leaf chunk in the tree, which in turn helps us to eliminate noises in a deep Web page. In this measure, remove the surplus noise and duplicate chunk using three parameters such as hyperlink percentage, Noise score and cosine similarity. Finally, obtain the main chunk extraction process using three parameters such as Title word Relevancy, Keyword frequency based chunk selection, Position features and set of keywords are extracted from those main chunks.

In the second phase,

- By using Fuzzy c-means clustering (FCM), the set of keywords were clustered for all deep web pages.

VII. Definitions of Terms Used In the Proposed Approach

Definition (chunk C): Consider a deep web page DW_P is segmented by blocks. These each blocks are known as chunk.

For example the web page is represented as, $DW_P = C_1, C_2, C_3 \ldots C_n$,Where the main chunk , $C_1 = C_{1,1}, C_{1,2} \ldots C_{m,n}$.

Definition (Hyperlink (HL_p)): A hyperlink has an anchor, which is the location within a document from which the hyperlink can be followed; the document having a hyperlink is called as its source document to web pages.

Hyperlink percentage $HL_P = \frac{n_l}{N}$

Where,

N → Number of Keywords in a chunk

n_l → Number of Link Keywords in a chunk

Definition (Noise score (N_s)): Noise score is defined as the ratio of number of images to total number of chunks.

Noise score, $N_s = \frac{n_I}{N_B}$

Where, n_I → Number of images in a chunk

N_B → Total number of images

Definition (Cosine similarity): Cosine similarity means calculating the similarity of two chunks. The inner product of the two vectors i.e., sum of the pairwise multiplied elements, is divided by the product of their vector lengths.

Cosine Similarity, $SIM_c C_1, C_2 = \frac{|C_1 . C_2|}{|C_1| \times |C_2|}$ where

C_1, C_2 →Weight of keywords in C_1, C_2

Definition (Position feature): Position features (PFs) that indicate the location of the data region on a deep web page. To compute the position feature score, the ratio (T) is computed and then, the following equation is used to find the score for the chunk.

$$PF_r = \begin{cases} 1 & 0.7 \geq T \\ 0 & Otherwise \end{cases} \quad (4)$$

Where,

T →

$$\frac{Number\ of\ keywords\ in\ Dtata\ \text{Re}gion\ chunk}{Number\ of\ keywords\ in\ Whole\ web\ page}$$

PF_r → Position features

Definition (Title word relevancy): A web page title is the name or heading of a Web site or a Web page. If there is more number of title words in a certain block, then it means that the corresponding block is of more importance.

Title word relevancy, $T_K = 1 - \left[\frac{m_k}{\left(m_k + \sum_{i=1}^{|m_k|} F(m_k^{(i)}) \right)} \right]$

Where,

m_k → Number of Title Keywords

$F(m_k^{(i)})$ → Frequency of the title keyword m_k in a chunk

Definition (Keyword frequency): Keyword frequency is the number of times the keyword phrase appears on a deep Web page chunk relative to the total number of words on the deep web page.

Keyword frequency based chunk selection, $K_f = \sum_{k=1}^{K} \frac{f_k}{N}$

Where,

f_k → Frequency of top ten keywords

N → Number of keywords

k → Number of Top-K Keywords

VIII. Proposed Approach to Vision-Based Deep Web Data Extraction for Web Document Clustering

Information extraction from web pages is an active research area. Recently, web information extraction has become more challenging due to the complexity and the diversity of web structures and representation. This is an expectable phenomenon since the Internet has been so popular and there are now many types of web contents, including text, videos, images, speeches, or flashes. The HTML structure of a web document has also become more complicated, making it harder to extract the target content. Until now, a large number of techniques have been proposed to address this problem, but all of them have inherent limitations because they are Web-page-programming-language dependent. In this paper, we present new approach for detection and removal of noisy data to

IX. Block Diagram

In the first phase, we are mainly concentrating to remove the following noises in stages: (1) Navigation bars, Panels and Frames, Page Headers and Footers,

Copyright and Privacy Notices, Advertisements and Other Uninteresting Data. (2) Duplicate Contents and (3) Unimportant Contents according to chunk importance. The removal of these noises is done by performing three operations. Firstly, using the chunk segmentation process, the noises such as the advertisements, images, audio, video, multiple links etc. are removed and only the useful text contents are segmented into chunks. Secondly, using three parameters such as *hyperlink percentage, Noise score* and *cosine similarity*, the surplus noise and duplicate chunks are removed to obtain the noiseless sub-chunks. And lastly, for each noiseless sub-chunk, we considered three parameters such as *Title word Relevancy ,Keyword frequency based chunk selection,* and *Position features*, using which we calculated the Sub-chunk weightage of each and every chunk. The high importance of the sub-chunks weightage consider as main-chunk weightage and the keywords are extracted from those main chunk. In the second phase, the set of keywords extracted are subjected to Fuzzy c-means clustering (FCM). The system model of the proposed technique which is extracting the important chunks and deep web clustering is shown schematically in Fig 1.

a) Phase 1: Vision-Based Deep Web Data Extraction

i. *Deep Web Page Extraction*

The Deep web is usually defined as the content on the Web not accessible through a search on general search engines. This content is sometimes also referred to as the hidden or invisible web. The Web is a complex entity that contains information from a variety of source types and includes an evolving mix of different file types and media. It is much more than static, self-contained Web pages. In our work, the deep web pages are collected from Complete Planet (www.completeplanet.com), which is currently the largest deep web repository with more than 70,000 entries of web databases.

ii. *Chunk Segmentation*

Web pages are constructed not only main contents information like product information in shopping domain, job information in a job domain but also advertisements bar, static content like navigation panels, copyright sections, etc. In many web pages, the main content information exists in the middle chunk and the rest of page contains advertisements, navigation links, and privacy statements as noisy data. Removing these noises will help in improving the mining of web. To assign importance to a region in a web page (W_P), we first need to segment a web page into a set of chunks.

extract main content information and deep web clustering that is both fast and accurate. The two phases and its sub-steps are given as follows.

- **Phase 1:** Vision-based deep web data identification
  - Deep web page extraction
  - Chunk segmentation
  - Noisy chunk Removal
  - Extraction of main chunk using chunk weightage
- **Phase 2:** Web document clustering
  - Clustering process using FCM

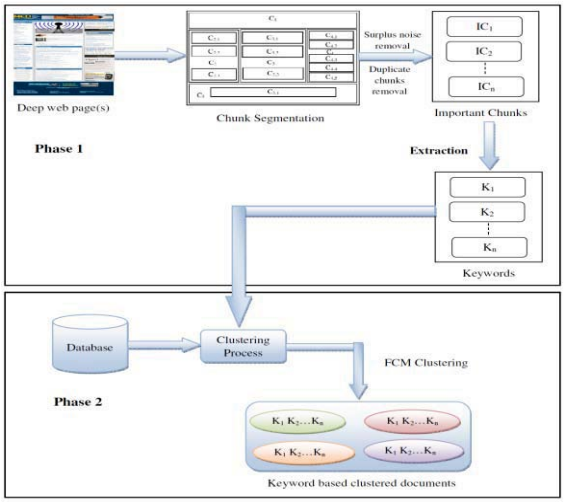


Fig. 1. Proposed method for extracting the important chunks and web clustering

Normally, a tag separated by many sub tags based on the content of the deep web page. If there is no tag in the sub tag, the last tag is consider as leaf node. The Chunk Splitting Process aims at cleaning the local noises by considering only the main content of a web page enclosed in div tag. The main contents are segmented into various chunks. The resultant of this process can be represented as follows:

$$C = \{C_1, C_2, C_3, \cdots, C_n\},\ C \in DW_P$$

Where, C → A set of chunks in the deep web page DW_P

n → Number of chunks in a deep web page DW_P

In fig.1 we have taken an example of a tree sample which consists of main chunks and sub chunks. The main chunks are segmented into chunks C_1, C_2 and C_3 using Chunk Splitting Operation and sub-chunks are segmented into $C_{2,1}, C_{2,2} \ldots C_{5,1}$ in fig 2.

iii. *Noisy Chunk Removal*

Surplus Noise Removal: A deep web page W_P usually contains main content chunks and noise chunks. Only the main content chunks represent the informative part that most users are interested in. Although other chunks are helpful in enriching functionality and guiding browsing, they negatively affect such web mining tasks as web page clustering and classification by reducing the accuracy of mined results as well as speed of processing. Thus, these chunks are called noise chunks. Removing these chunks in our research work,

we have concentrated on two parameters; they are Hyperlink Percentage (HL_p) and Noise score (N_s) which is very significant. The main objective for removing noise from a Web Page is to improve the performance of the search engine.

The representation of each parameter is as follows:

1. *Hyperlink Keyword* (HL_p) – A hyperlink has an anchor, which is the location within a document from which the hyperlink can be followed; the document containing a hyperlink is known as its source document to web pages. Hyperlink Keywords are the keywords which are present in a chunk such that it directs to another page. If there are more links in a particular chunk then it means the corresponding chunk has less importance. The parameter Hyperlink Keyword Retrieval calculates the percentage of all the hyperlink keywords present in a chunk and is computed using following equation.

$$\textit{Hyperlink word Percentage, } HL_P = \frac{n_l}{N}$$

Where,

N → Number of Keywords in a chunk

n_l → Number of Link Keywords in a chunk

2. *Noise score* (N_s) – The information on Web page W_P consists of both texts and images (static pictures, flash, video, etc.). Many Internet sites draw income from third-party advertisements, usually in the form of images sprinkled throughout the site's pages. In our work, the parameter Noise score calculates the percentage of all the images present in a chunk and is computed using following equation. $\textit{Noise score, } N_s = \frac{n_I}{N_B}$

Where,

n_I → Number of images in a chunk

N_B → Total number of images

Duplicate Chunk Removal Using Cosine Similarity: *Cosine Similarity*. Cosine similarity is one of the most popular similarity measure applied to text documents, such as in numerous information retrieval applications [7] and clustering too [8]. Here, duplication detection among the chunk is done with the help of cosine similarity.

Given two chunks C_1 and C_2, their cosine similarity is

Cosine Similarity $SIM_c(C_1, C_2) = \frac{|C_1.C_2|}{|C_1| \times |C_2|}$

Where,

C_1 , C_2 →Weight of keywords in C_1, C_2

iv. *Extraction of Main Chunk*

Chunk Weightage for Sub-Chunk: In the previous step, we obtained a set of chunks after removing the noise chunks and duplicate chunks present in a deep web page. Web page designers tend to organize their content in a reasonable way: giving prominence to important things and deemphasizing the unimportant parts with proper features such as position, size, color, word, image, link, etc. A chunk importance model is a function to map from features to importance for each chunk, and can be formalized as: $\langle chunk\ \ features \rangle \Rightarrow chunk\ \text{importance}$

The preprocessing for computation is to extract essential keywords for the calculation of Chunk Importance. Many researchers have given importance to different information inside a webpage for instance location, position, occupied area, content, etc. In our research work, we have concentrated on the three parameters Title word relevancy, keyword frequency based chunk selection, and position features which are very significant. Each parameter has its own significance for calculating sub-chunk weightage. The following equation computes the sub-chunk weightage of all noiseless chunks.

$$C_w = \alpha T_k + \beta\ K_f + \ \gamma\ PF_r \qquad (1)$$

Where α, β, γ → Constants

For each noiseless chunk, we have to calculate these unknown parameters T_K , K_f and PF_r . The representation of each parameter is as follows:

1. *Title Keyword* – Primarily, a web page title is the name or title of a Web site or a Web page. If there is more number of title words in a particular block then it means the corresponding block is of more importance. This parameter Title Keyword calculates the percentage of all the title keywords present in a block. It is computed using following equation.

Title word Relevancy, $T_K = 1 - \left[\frac{m_k}{\left(m_k + \sum_{i=1}^{|m_k|} F(m_k^{(i)}) \right)} \right]$ (2)

Where, m_k → Number of Title Keywords

T_k → Title word relevancy, $F(m_k^{(i)})$ → Frequency of the title keyword n_t in a chunk

2. *Keyword Frequency based chunk selection:* Basically, Keyword frequency is the number of times the keyword phrase appears on a deep Web page chunk relative to the total number of words on the deep web page. In our work, the top-K keywords of each and every chunk were selected and then their frequencies were calculated. The parameter

keyword frequency based chunk selection calculates for all sub-chunks and is computed using following equation.

Keyword Frequency based chunk selection

$$K_f = \sum_{k=1}^{K} \frac{f_k}{N} \quad (3)$$

Where,

f_k → Frequency of top ten keywords

K_f → Keyword Frequency based chunk selection

k → Number of Top-K Keywords

3. *Position features (PFs):* Generally, these data regions are always centered horizontally and for calculating, we need the ratio (T) of the size of the data region to the size of whole deep Web page instead of the actual size. In our experiments, the threshold of the ratio is set at 0.7, that is, if the ratio of the horizontally centered region is greater than or equal to 0.7, then the region is recognized as the data region. The parameter position features calculates the important sub chunk from all sub chunk and is computed using following equation.

$$PF_r = \begin{cases} 1 & 0.7 \geq T \\ 0 & Otherwise \end{cases} \quad (4)$$

Where,

$T \rightarrow \frac{Number\ of\ keywords\ in\ Dtata\ Region\ chunk}{Number\ of\ keywords\ in\ Whole\ web\ page}$

PF_r → Position features

Thus, we have obtained the values of T_K, K_f and PF_r by substituting the above mentioned equation. By substituting the values of T_K, K_f and PF_r in eq.1, we obtain the sub-chunk weightage.

Chunk Weightage for Main Chunk: We have obtained sub-chunk weightage of all noiseless chunks from the above process. Then, the main chunks weightage are selected from the following equation

$$C_i = \sum_{i=1}^{n} \alpha\ c_w^{(i)} \quad (5)$$

Where, $c_w^{(i)}$ → i^{th} Sub-chunk weightage of Main-chunk. α → Constant, C_i → Main chunk weightage

Thus, finally we obtain a set of important chunks and we extract the keywords from the above obtained important chunks for effective web document clustering mining.

Input : Deep web page W_p , Input Constants α, β, γ

Output : Set of cluster with relevant keywords CS_k

Pseudo code

1. Input the deep web pages, W_P
2. Deep web page segmentation using tag
3. Compute Noise chunk removal value for each leaf nodes of deep web pages

3.1 Compute surplus noise removal for all leaf nodes using Hyperlink Percentage and noise score.

3.1.1 Compute Hyperlink word Percentage, $HL_P = \frac{n_l}{N}$

3.1.2 Compute Noise score, $N_s = \frac{n_I}{N_B}$

3.2 Compute duplicate noise removal for all leaf nodes using cosine similarity

$$SIM_c C_1, C_2 = \frac{|C_1.C_2|}{|C_1| \times |C_2|}$$

4. Compute sub-chunk weightage value SC_w for each Noiseless chunk of deep Web pages

4.1 Compute Title word relevancy for noiseless chunk

$$T_K = 1 - \left[\frac{n_t}{\left(n_t + \sum_{i=1}^{|n_t|} F(n_t^{(i)}) \right)} \right]$$

4.2 Compute Keyword frequency based chunk importance

$$K_f = \frac{\sum_{k=1}^{TopK} f_k}{N}$$

4.3 Compute Position features based chunk importance

$$PF_r = \begin{cases} 1 & 0.7 \geq T \\ 0 & Otherwise \end{cases}$$

4.4 Compute sub-chunk weightage

$$SC_w = \alpha H_f + \beta F_f + \gamma PF_r$$

5. Compute main-chunk weightage value M_i , i.e. set of keywords

$$M_i = \sum_{i=1}^{n} \alpha \, e_w^{(i)}$$

6. After computing the extraction of keywords for all deep web pages, set of

Keywords were clustered using Fuzzy c-means clustering.

$$CS_k = \{ CS_{k1}, CS_{k2} ... CS_{kn} \}$$

b) Phase Ii: Deep Web Documnet Clustering Using Fcm

Let DB be a dataset of web documents, where the set of keywords is denoted by $k = \{k_1, k_2, ..., k_n\}$. Let $X = \{x_1, x_2, ..., x_N\}$ be the set of N web documents, where $x_i = \{x_{i1}, x_{i2}, ..., x_{in}\}$.Each $x_{ij} (i = 1,..., N; j = 1,..., n)$ corresponds to the frequency of keyword x_i on web document. Fuzzy c-means [29] partitions set of N web documents in R^d dimensional space into c $(1< c <n)$ fuzzy clusters with $Z = \{z_1, z_2, ..., z_c\}$ cluster centers or centroids. The fuzzy clustering of keywords is described by a fuzzy matrix μ with n rows and c columns in which n is the number of keywords and c is the number of clusters. μ_{ij}, the element in the i^{th} row and j^{th} column in μ, indicates the degree of association or membership function of the i^{th} object with the j^{th} cluster. The characters of μ are as follows:

$$\mu_{i,j} \in [0,1] \quad \forall i = 1,2,...,n; \quad \forall j = 1,2,...,c; \tag{6}$$

$$\sum_{j=1}^{c} \mu_{ij} = 1 \quad \forall i = 1,2,...,n; \tag{7}$$

$$0 < \sum_{i=1}^{n} \mu_{ij} < n \quad \forall j = 1,2,...,c; \tag{8}$$

The objective function of FCM algorithm is to minimize the Eq. (9):

$$J_m = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^m d_{ij} \tag{9}$$

Where

$$d_{ij} = \| k_i - z_j \| \tag{10}$$

in which, m(m >1) is a scalar termed the weighting exponent and controls the fuzziness of the resulting clusters and d_{ij} is the Euclidian distance from k_i to the cluster center z_i .The z_j ,centroid of the j^{th} cluster, is obtained using Eq. (11)

$$z_j = \frac{\sum_{i=1}^{n} \mu_{ij}^m k_i}{\sum_{i=1}^{n} \mu_{ij}^m} \tag{11}$$

The FCM algorithm is iterative and can be stated as follows

Algorithm 2.Fuzzy c-means:

1. Select m (m > 1); initialize the membership function values μ_{ij} , $i = 1,2,...,n$; $j = 1,2,...,c.$
2. Compute the cluster centers z_j , $j = 1,2,...,c,$ according to Eq. (11).
3. Compute Euclidian distance d_{ij} , $i = 1,2,...,n$; $j = 1,2,...,c.$
4. Update the membership function μ_{ij} , $i = 1,2,...,n; j = 1,2,...,c$ according to Eq. (12).

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m-1}}} \tag{12}$$

X. Results and Discussion

a) Experimental Set Up

The experimental results of the proposed method for vision-based deep web data extraction for web document clustering are presented in this section. The proposed approach has been implemented in java (jdk 1.6) and the experimentation is performed on a 3.0 GHz Pentium PC machine with 2 GB main memory. For experimentation, we have taken many deep web pages which contained all the noises such as Navigation bars, Panels and Frames, Page Headers and Footers, Copyright and Privacy Notices, Advertisements and Other Uninteresting Data. These pages are then applied to the proposed method for removing the different

noises. The removal of noise blocks and extracting of useful content chunks are explained in this sub-section. Finally, extracting the useful content keywords are clustered using Fuzzy c-means clustering.

b) Data Sets

GDS: Our data set is collected from the complete planet web site (www.completeplanet.com). Complete-planet is currently the largest depository for deep web, which has collected the search entries of more than 70,000 web databases and search engines. These Web databases are classified into 42 categories covering most domains in the real world. GDS contains 1,000 available Web databases. For each Web database, we submit five queries and gather five deep Web pages with each containing at least three data records. SDS: Special data set (SDS). During the process of obtaining GDS, we noticed that the data records from two-thirds of the Web databases have less than five data items on average. To test the robustness of our approaches, we select 100 Web databases whose data records contain more than 10 data items from GDS as SDS.

Experimental results

c) Performance Analysis of Phase 2 of Our Technique

While analyzing the results of GDS and SDS datasets, the accuracy, computation time and the memory usage is evaluated in clustering process. Accuracy: The accuracy values obtained for two different datasets are plotted in the figure 8, in which the dataset 1 and dataset 2 are achieves better accuracy (40%) in Fig 8. Execution Time: The run time performance of the methods is plotted as a graph shown in Fig 9, in which the dataset 2 achieves better execution time (0.906 sec) compared with data set 1. Memory usage: By analyzing the figure 10, the dataset 1 are achieves more memory usage (2990 KB) compared with dataset 2.

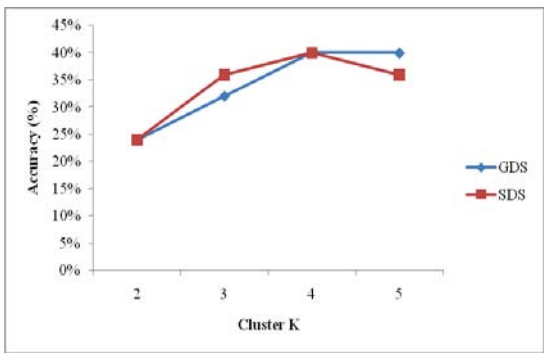


Figure 2 : clustering accuracy of dataset 1(GDS) and data set 2(SDS)

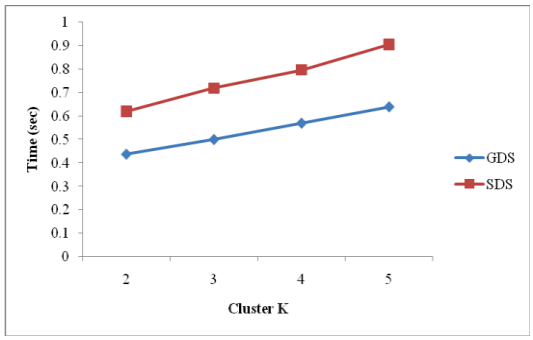


Figure 3 : Time of dataset 1(GDS) and data set 2(SDS)

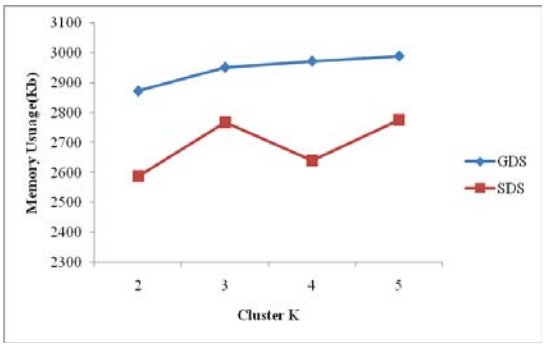


Figure 4 : Memory usage of dataset 1(GDS) and dataset 2(SDS)

XI. Conclusion

In this paper, an approach to vision-based deep web data extraction is proposed for web document clustering. The proposed approach comprises of two phases: 1) Vision-based web data extraction and 2) web document clustering. In phase 1, the web page information is classified into various chunks. From which, surplus noise and duplicate chunks are removed using three parameters, such as hyperlink percentage, noise score and cosine similarity. To identify the relevant chunk, three parameters such as Title word Relevancy, Keyword frequency-based chunk selection, Position features are used and then, a set of keywords are extracted from those main chunks. Finally, the extracted keywords are subjected to web document clustering using Fuzzy c-means clustering (FCM). Our experimental results showed that the proposed VDEC method can achieve stable and good results for both datasets.

References Références Referencias

1. P S Hiremath, Siddu P Algur,"Extraction of data from web pages: a vision based approach," International Journal of Computer and Information Science and Engineering, Vol.3, pp.50-59, 2009.
2. Jing Li, "Cleaning Web Pages for Effective Web Content Mining, "In Proceedings: DEXA, 2006.
3. Thanda Htwe,"Cleaning Various Noise Patterns in Web Pages for Web Data Extraction," International

Journal of Network and Mobile Technologies,vol.1,no.2,2010.
4. Yang, Y. and Zhang, H., "HTML Page Analysis Based on Visual Cues," In 6th International Conference on Document Analysis and Recognition, Seattle, Washington, USA, 2001.
5. Longzhuang Li, Yonghuai Liu, Abel Obregon," Visual Segmentation-Based Data Record Extraction from Web Documents,"IEEE International Conference on Information Reuse and Integration, pp.502 – 507, 2007.
6. Qingshui Li; Kai Wu; "Study of Web Page Information topic extraction technology based on vision,"IEEE International Conference on Computer Science and Information Technology (ICCSIT), vol.9, pp.781-784, 2010.
7. R. B. Yates and B. R. Neto,"Modern Information Retrieval," Addison-Wesley, New York, 1999.
8. B. Larsen and C. Aone. "Fast and effective text mining using linear-time document clustering," In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1999.
9. Chen Hong-ping; Fang Wei; Yang Zhou; Zhuo Lin; Cui Zhi-Ming; "Automatic Data Records Extraction from List Page in Deep Web Sources, "Asia-Pacific Conference on Information Processingvol.1,pp.370-373, 2009.
10. Zhang Pei-ying, Li Cun-he,"Automatic text summarization based on sentences clustering and extraction,"2nd IEEE International Conference on Computer Science and Information Technology, pp.167-170, 2009.
11. Yan Guo, Huifeng Tang, Linhai Song, Yu Wang, Guodong Ding, "ECON: An Approach to Extract Content from Web News Page", In Proceedings of the 12th International Asia-Pacific Web Conference (APWEB), pp. 314 – 320, April 06-April 08 Buscan, Korea, 2010
12. Wei Liu, Xiaofeng Meng, Weiyi Meng,"ViDE: A Vision-Based Approach for Deep Web Data Extraction,"IEEE Transactions on Knowledge and Data Engineering, vol.22, no.3, pp.447-460, 2010.
13. Ashraf, F.; Ozyer, T.; Alhajj, R.; "Employing Clustering Techniques for Automatic Information Extraction from HTML Documents,"IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol.38, no.5, pp.660-673, 2008.
14. Manisha Marathe, Dr. S.H.Patil, G.V.Garje,M.S.Bewoor, "Extracting Content Blocks from Web Pages", International Journal of Recent Trends in Engineering, Vol .2, No. 4, November 2009.
15. Sandip Debnath,Prasenjit Mitra,C. Lee Giles, "Automatic Extraction of Informative Blocks from WebPages", In Proceedings of the ACM symposium on Applied computing, Santa Fe, New Mexico, pp. 1722 – 1726,2005.
16. Lan Yi ,Bing Liu, "Web page cleaning for web mining through feature weighting", In Proceedings of the 18th international joint conference on Artificial intelligence, pp. 43-48 , August 09 - 15 ,Acapulco, Mexico, 2003
17. K. Tripathy , A. K. Singh , "An Efficient Method of Eliminating Noisy Information in Web Pages for Data Mining", In Proceedings of the Fourth International Conference on Computer and Information Technology, pp. 978 – 985, 2004.
18. Zhao Cheng-li and Yi Dong-yun, "A method of eliminating noises in Web pages by style tree model and its applications", Wuhan University Journal of Natural Sciences, Wuhan University, co-published with Springer Vol.9, No.5, pp. 611-616, 2004.
19. Ruihua Song,Haifeng Liu,Ji-Rong Wen, Wei-Ying Ma, "Learning Block Importance Models for Web Pages", Proceedings of the 13th international conference on World Wide Web, pp. 203 - 211 , New York, NY, USA, 2004.
20. Ruihua Song, Haifeng Liu, Ji-Rong Wen, Wei-Ying Ma, "Learning Important Models for Web Page Blocks based on Layout and Content Analysis", ACM SIGKDD Explorations Newsletter, Vol. 6 , No. 2,pp. 14 - 23 ,2004. 2-57, 1973.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 5 Version 1.0 March 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

An Enhanced Scheduling Algorithm for Qos Optimization in 802.11e Based Networks

By Ms.V.R.Azhaguramyaa, S.J.K.Jagadeesh Kumar & P.Parthasarathi
Sri Krishna College of Technology

Abstract - Quality of Service (QoS) is the ability to guarantee a certain level of performance to a data flow ie., guaranteeing required bit rate, delay, etc. IEEE 802.11 a/b/g networks do not provide QoS differentiation among multimedia traffic. QoS provisioning is one of the essential features in IEEE 802.11e. It uses Enhanced Distributed Channel Access (EDCA) which is a contention-based channel access mode to provide QoS differentiation. EDCA works with four Access Categories (AC). Differentiation of Access Categories are achieved by differentiating the Arbitration Inter-Frame Space (AIFS), the initial contention window size (CWmin), the maximum contention window size (CWmax) and the transmission opportunity (TXOP). However AIFS, CWmin, CWmax are considered to be fixed for a given AC, while TXOP may be varied. A TXOP is a time period when a station has the right to initiate transmissions onto the wireless medium. By varying the TXOP value among the ACs the QoS optimization- throughput stability and minimum delay is achieved. EDCA has many advantages such as it fully utilizes the channel bandwidth, and does not require centralized admission control and scheduling algorithms over the contention-free access mode.

Keywords : *EDCA, MAC, IEEE 802.11e, Quality of Service, QoS optimization*

GJCST Classification: *C.2.1*

Strictly as per the compliance and regulations of:

An Enhanced Scheduling Algorithm for Qos Optimization in 802.11e Based Networks

Ms.V.R.Azhaguramyaa [α], Prof.S.J.K.Jagadeesh Kumar [σ] & P.Parthasarathi [ρ]

Abstract - Quality of Service (QoS) is the ability to guarantee a certain level of performance to a data flow ie., guaranteeing required bit rate, delay, etc. IEEE 802.11 a/b/g networks do not provide QoS differentiation among multimedia traffic. QoS provisioning is one of the essential features in IEEE 802.11e. It uses Enhanced Distributed Channel Access (EDCA) which is a contention-based channel access mode to provide QoS differentiation. EDCA works with four Access Categories (AC). Differentiation of Access Categories are achieved by differentiating the Arbitration Inter-Frame Space (AIFS), the initial contention window size (CWmin), the maximum contention window size (CWmax) and the transmission opportunity (TXOP). However AIFS, CWmin, CWmax are considered to be fixed for a given AC, while TXOP may be varied. A TXOP is a time period when a station has the right to initiate transmissions onto the wireless medium. By varying the TXOP value among the ACs the QoS optimization- throughput stability and minimum delay is achieved. EDCA has many advantages such as it fully utilizes the channel bandwidth, and does not require centralized admission control and scheduling algorithms over the contention-free access mode.

IndexTerms : EDCA, MAC, IEEE 802.11e, Quality of Service, QoS optimization

I. Introduction

a) Background

There have been various Frequency-based approaches are available for QoS optimization but they incur high computational complexity because modeling the AIFS, CWmin and CWmax values require solving non-linear equation systems that are extremely computationally demanding and not suitable for real-time applications. QoS optimization in contention-free mode requires centralized admission and scheduling algorithms, thus not flexible The Enhanced Distributed Channel Access mode is able to provide QoS optimization by easily controlling the TXOP.

b) IEEE 802.11e

IEEE 802.11e-2005 or 802.11e [3] is an approved amendment to the IEEE 802.11 standard that defines a set of Quality of Service enhancements for wireless LAN applications through modifications to the Media Access Control (MAC) layer.

Author [α]: PG Student, Dept. of CSE Sri Krishna College of Technology Coimbatore, India E-mail : vrazhaguramyaa@gmail.com
Author [σ]:Professor & Head, Dept. of CSE Sri Krishna College of Technology Coimbatore, India E-mail :jagadeesh_sk@rediffmail.com
Author [ρ]:Asst. Professor, Dept. of CSE Sri Krishna College of Technology Coimbatore, India E-mail :sarathi.pp@gmail.com

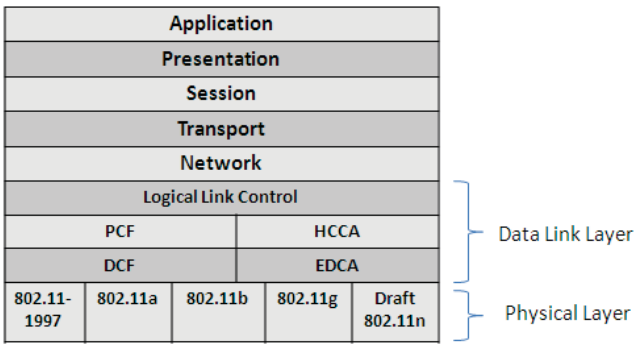


Fig. 1 OSI model of IEEE 802.11e

c) 802.11e MAC Operation

The 802.11e enhances the DCF and the PCF, through a new coordination function: the hybrid coordination function (HCF). Within the HCF, there are two methods of channel access, similar to those defined in the legacy 802.11 MAC: HCF Controlled Channel Access (HCCA) and Enhanced Distributed Channel Access (EDCA) which is illustrated in Fig. 1.Both EDCA and HCCA define Access Categories.

d) Importance of Access Categories in 802.11e

The 802.11e MAC supports the access categories which are listed in Table I.

Table I
Access categories

ACCESS CATEGORIES	DESCRIPTION
AC_VO(Voice)	Voice traffic and network control belong to AC_VO.
AC_VI(Video)	Video and video/controlled load belong to AC_VI.
AC_BE(Best Effort)	Best effort (E-mail) and video/excellent effort belong to AC_BE.
AC_BK(Background)	Background (Uninvited) traffic will come under AC_BK.

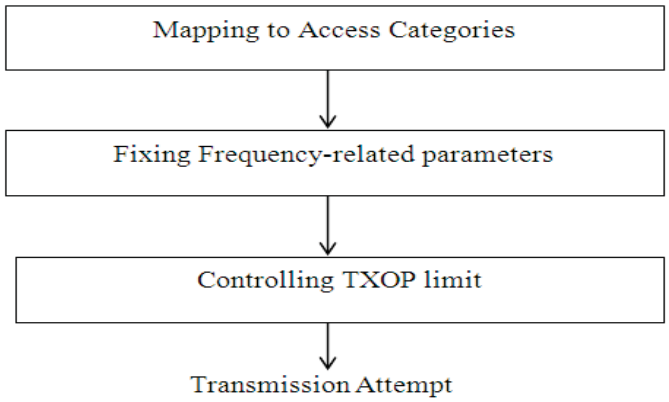


Fig. 2 QoS optimization process

e) Qos Optimization

Fig. 2 describes the process of QoS optimization. Initially heterogeneous traffic reaches the MAC and they are mapped to the corresponding Access Categories. Then all frequency-related parameters of various Access Categories are fixed, by controlling the TXOP Limit parameter the higher priority traffic has a higher chance of being sent and waits a little less before it sends its packet, on average, than a station with low priority traffic.

II. Related Work

Bellalta. B.*et al* investigated the basic values of EDCA parameters which should be changed to perform the QoS optimization [1]. Their work is mainly based on the frequency of acquiring transmission opportunities parameters and they have concentrated, to maximize the elastic (BE) throughput while assuring the bandwidth-delay requirements of the rigid flows(VO).

Zhen-ning, Kong *et al.* analyzed the performance of contention-based channel access in IEEE 802.11e [7] and they have produced the markov chain model of one Access Category per station. They have concentrated on AIFS in the Enhanced Distributed Channel Access.

In this paper we propose a new optimization algorithm which is the modification of EDCA and that new algorithm provides per stream QoS which is not available in EDCA [2] and it is achieved by tuning the duration of transmission opportunity parameter called TXOP limit.

This new work follows the implementation details outlined by, Khaled A. Shuaib. The author specified the cell structure of wireless networks, important parameters needed for creating the simula used in Qualnet simulation tool [5].

III. Proposed Scheme

a) EDCA

The Enhanced Distributed Channel Access (EDCA) mechanism of 802.11e extends the basic 802.11 DCF algorithm with Quality of Service capabilities. In EDCA mode, packets are categorized into prioritized classes, called access categories (ACs). In EDCA mode, airtime allocation among traffic sessions in different ACs is differentiated by assigning each AC with different EDCA parameters. Differential allocation of airtime to different ACs is essential for QoS-enabled applications.

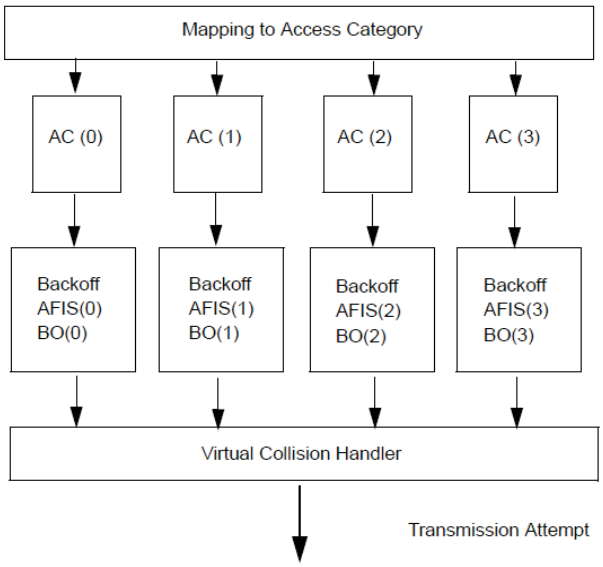


Fig. 3 EDCA

b) Edca Parameters

There are two sets of EDCA parameters that can achieve airtime differentiation.

1) Frequency of transmission opportunities: The frequency of transmission opportunities is determined through three parameters:

- Arbitration Inter Frame Space (*AIFS*)
- Minimum Contention Window Size (*CWmin*)
- Maximum Contention Window Size (*CWmax*)

Each AC maintains a contention window variable (CW), which is initialized to CWmin. The CW is incremented after transmission failures until it reaches CWmax, and is reset to CWmin after a successful transmission. To avoid collisions, a backoff timer is independently chosen from the range [0, CW] for each AC. Since smaller CWmin and CWmax generally lead to smaller CW values, they result in shorter backoff timer and higher transmission opportunity frequency.

2) Duration of transmission opportunity: The maximum allowed duration for each acquired transmission opportunity is determined by a parameter called, - *TXOP limit*.

Once a station acquires a transmission opportunity, it may transmit multiple frames within the assigned *TXOP* limit. Assigning different *TXOP* values to ACs, therefore, achieves differential airtime allocations

[2]. Controlling the TXOP limit allows us to derive a simple, closed-form equation for the effective airtime.

c) EDCA Mechanism

A new concept, transmission opportunity (TXOP), is introduced in IEEE 802.11e. A TXOP is a time period when a station has the right to initiate transmissions onto the wireless medium. A station cannot transmit a frame that extends beyond a TXOP. EDCA works with four Access Categories (ACs), as shown in Fig. 3, this differentiation is achieved through varying the amount of time; a station would sense the channel to be idle, and the length of the contention window for a backoff. Differentiated ACs is achieved by differentiating AIFS, the initial window size and the maximum window size. EDCA employs AIFS[i], CWmin[i], and CWmax[i] (all for i = 0... 3) instead of DIFS, minCW and maxCW, respectively [4].

d) QOS Optimization in IEEE 802.11e Using EDCA

Table III

Values of EDCA parameters

AC	AIFSN	TXOP (ms)	CWmin	CWmax
AC[0] - BK	7	0	CWmin	CWmax
AC[1] - BE	3	0	CWmin	CWmax
AC[2] - VI	2	6.016	CWmin/2	CWmin
AC[3] - VO	2	3.264	CWmin/4	CWmin/2

Each frame arriving at the MAC with a priority is mapped into an AC. The Table III specifies the values of EDCA parameters which belong to different access categories of IEEE 802.11e. AIFS for a given AC is determined by the following equation:

$$AIFS = SIFS + AIFSN * aSlotTime \quad (1)$$

Where AIFSN is AIFS Number and aSlotTime is the duration of a time slot [1]. The AC with the smallest AIFS has the highest priority. The CWmin value is 31 and the CWmax value is 1023.

e) TXOP LIMIT- The Controlling Knob

Consider S wireless stations compete for the shared air medium of a wireless LAN using the IEEE 802.11e EDCA protocol. These wireless stations transmit data to/from the base station at different bit rates, and the rate differentiation is achieved by varying the TXOP limits for individual wireless stations. In optimization problem, it is a need to determine the total effective airtime (EA) of the wireless medium so that it can be divided among stations, and to avoid over/under allocation of the wireless medium. The virtual transmission time v_j as the time duration between the j-th and the (j + 1)-th successful transmissions is defined.

Each virtual transmission consists of three periods:

- Idle
- Collision
- Transmission

Let consider $E[x]$ to denote the average transmission opportunity limit for all wireless stations, and $E[v]$ to denote the average virtual transmission time. Then, the effective airtime can be given by:

$$EA = E[x]/E[v] \quad (2)$$

Let denote the number of collisions in a virtual transmission time by C, define i_k to be the duration of the k-th idle period, and similarly, c_k to be the duration of the k-th collision period. Then $E[v]$ is given by:

$$E[v] = E[C](E[c] + t_d + t_s + t_a) + (E[C] + 1) E[I] + E[x] + t_d \quad (3)$$

Where,

t_d is the distributed inter-frame space (DIFS),

t_s is the short inter-frame space (SIFS),

t_a is the average time of sending an acknowledgment.

From the equation (3) it is found that optimal solution for airtime differentiation comes from controlling the *TXOP* limit and by fixing the frequency of transmissions opportunities parameters.

IV. Experimental Results

Scenario consists of 1 Access Point and 24 wireless stations. Fig. 5 shows the simulation of created scenario and its progress. Scenario has the following properties [5]:

Simulation tool –QualNet 5.0.2
PHY Layer Model – IEEE 802.11b
Bandwidth – 20MHZ (min)
Antenna Height (PHY) – 1.5 meters
Terrain Space – 500*500 meters
Simulation time – 100 seconds
Scheduling algorithm used to select service classes – Strict Priority
Scheduling algorithm used inside the service class- FIFO.

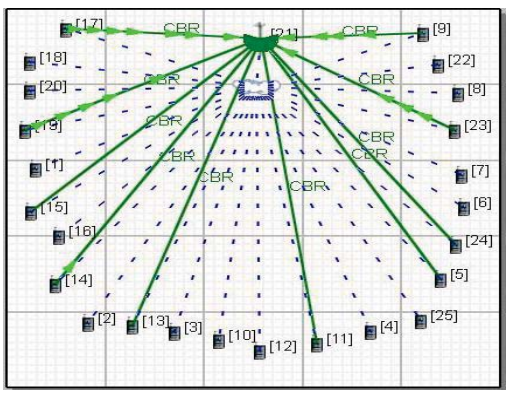


Fig. 5 Scenario Simulation

a) Throughput analysis

Throughput is calculated for both Access point and wireless stations using the following formula:

Access Point

*Throughput= ((Total no. of bytes received * 8)/ Session Duration)*

Wireless Stations

*Throughput= ((Total no. of bytes sent * 8) / Session Duration)*

1) Access Point Throughput: Fig. 6 shows the throughput of server, the Access point node (21) is the server, throughput of 21 is very high because it always receives packets from the high priority traffic because of the TXOP variation. This result is taken from the statistics (.stat) file of created scenario.

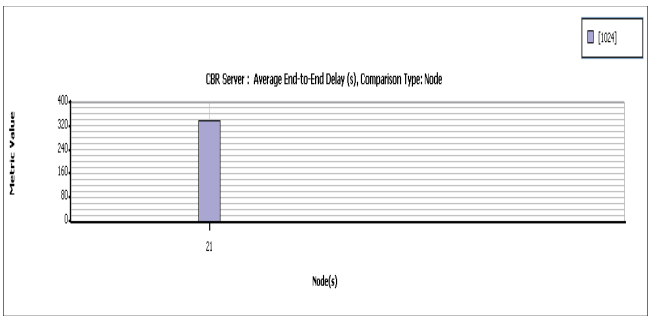


Fig. 6 Access point throughput analysis

2) Wireless stations Throughput: Fig. 7 shows the throughput of wireless stations which generates various traffic. The analysis says that the throughput of nodes which generates high priority traffic (9, 17, 19 and 23) is very high because of the TXOP variation.

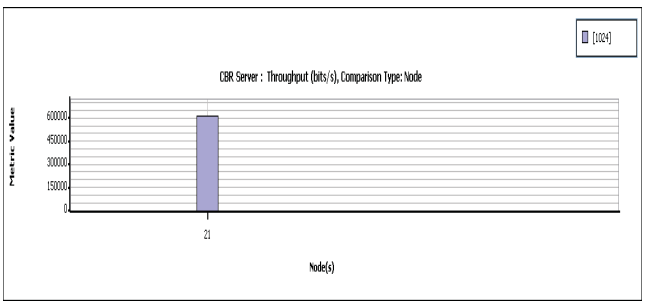


Fig. 7 Wireless stations throughput analysis

b) Average End-To-End Delay Analysis

Average end-to-end delay is calculated at the Access Point using the following formula.

CBR Avg. End-to-end delay = (Sum of the delays of each CBR packet received) / (no. of CBR packets received)

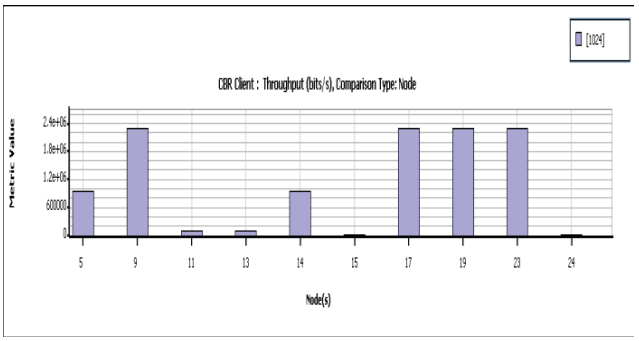


Fig. 8 Average End-to-End delay analysis

Fig. 8 shows the average end-to-end delay of the Access Point. From the statistics (.stat) file of created scenario, it is clear that the delay of high priority traffic is comparatively less than the low priority traffic.

V. Conclusion

The QoS optimization is provided by Enhanced Distributed Channel Access mode in IEEE 802.11e based Networks. With EDCA, packets are categorized into prioritized classes, higher priority traffic has a higher chance of being sent and waits a little less before it sends its packet, on average, than a station with low priority traffic. Using EDCA the quality improvement comes at negligible cost, because the optimal solution is computed using simple equations. EDCA is suited for networks which support link-layer traffic differentiation.

In future, the EDCA mechanism can be implemented for IEEE 802.16 based networks and the cross layering framework can also be included to improve the QoS optimization.

References Références Referencias

1. Bellalta. B., Cano. C., Oliver. M. and Meo. M., "Modeling the IEEE 802.11e EDCA for MAC Parameter Optimization", *Het-Nets 06,* Bradford, UK, September 2006.
2. Cheng-Hsin Hsu, Mohamed Hefeeda, "A Framework for Cross-Layer Optimization of Video Streaming in Wireless Networks", *ACM Transactions on Multimedia Computing, Communications and Applications,* Vol. 7, No. 1, Article 5,January 2011.
3. IEEE Std 802.11e; Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications; *Amendment: Medium Access Control (MAC) QoS Enhancements*, IEEE Std 802.11e-2005.
4. Jochen Schiller, "*Mobile Communications*", Second Edition, Pearson Education, 2003.
5. Khaled A. Shuaib, "A Performance Evaluation Study of WiMAX Using Qualnet" *Proceedings of the World Congress on Engineering 2009 Vol I* WCE 2009, July 1 - 3, 2009.

6. Y. C. Tay and K. C. Chua, "A capacity analysis for the IEEE 802.11 MAC protocol," *Wireless Networks*, vol. 7, pp. 159–171, July 2001.
7. Zhen-ning Kong Danny H. K. Tsang Brahim Bensaou and Deyun Gao,"Performance Analysis of IEEE 802.11e Contention-Based Channel Access", *IEEE Journal On Selected Areas In Communications,* Vol. 22, No. 10, December 2004.
8. http://www.qualnet.com

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 5 Version 1.0 March 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Efficient Image Retrieval Based On Texture Features Using Concept of Histogram

By Dr. Neelima Sadineni, DavarapalliAnandam, Venkata Ramesh Pokala, Pasupuletu Rajesh & KolakaluriSrinivasaRao
Anna University,Chennai

Abstract - Image retrieval is fast growing research oriented area now days. As information retrieval plays a major role in transmitting knowledge both in the forms of text and images, image retrieval got a major focus. In this paper we integrated the Histogram Intersection measure method to compare the query image with database images; by this approach we can measure over-all similarity between images, by incorporating all local properties of the texture histograms of the images through which we proved that our approach in retrieving the image is accurate.

Keywords : *Histogram, image, texture, database, retrieval*

GJCST Classification: *I.4.7, I.4.5*

Strictly as per the compliance and regulations of:

Efficient Image Retrieval Based On Texture Features Using Concept of Histogram

Dr. Neelima Sadineni [α],DavarapalliAnandam [σ],Venkata Ramesh Pokala [ρ],Pasupuletu Rajesh [ω] & KolakaluriSrinivasaRao[¥]

Abstract - Image retrieval is fast growing research oriented area now days. As information retrieval plays a major role in transmitting knowledge both in the forms of text and images, image retrieval got a major focus. In this paper we integrated the Histogram Intersection measure method to compare the query image with database images; by this approach we can measure over-all similarity between images, by incorporating all local properties of the texture histograms of the images through which we proved that our approach in retrieving the image is accurate.

Keywords : Histogram, image, texture, database, retrieval

I. Introduction

To date, image and video storage and retrieval systems have typically relied on human supplied textual annotations to enable indexing and searches. The text-based indexes for large image and video archives are time consuming to create. They necessitate that each image and video scene is analyzed manually by a domain expert so the contents can be described textually. The language-based descriptions, however, can never capture the visual content sufficiently. For example, a description of the overall semantic content of an image does not include an enumeration of all the objects and their characteristics, which may be of interest later. A content mismatch occurs when the information that the domain expert ascertains from an image differs from the information that the user is interested in. A content mismatch is catastrophic in the sense that little can be done to approximate or recover the omitted annotations. In addition, a language mismatch can occur when the user and the domain expert use different languages or phrases. Because text-based matching provides only hit-or-miss type searching, when the user does not specify the right keywords the desired images are unreachable without examining the entire collection.

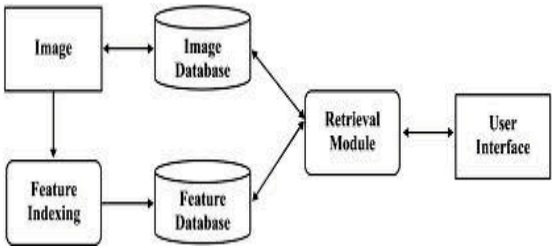


Fig.1: The Architecture of CBIR Technique

The prime requirement for Retrieval systems is to be able to display images relating to a named query image. The text indexing is often limited, tedious and subjective for describing image content. So there is increasing interest in the use of CBIR techniques. The problems with text-based access to images have prompted increasing interest in the development of image based solutions. This is more often referred to as Content Based Image Retrieval (CBIR) as shown in Fig.1. Content Based Image Retrieval relies on the characterization of primitive features such as color, shape and texture that can be automatically extracted from the images themselves. Queries to CBIR system are most often expressed as visual exemplars of the type of the image or image attributed being sought. For Example user may submit a sketch, click on the texture pallet, or select a particular shape of interest. This system then identifies those stored images with a high degree of similarity to the requested feature.

Digital imaging has become the standard for all image acquisition devices. So there is an increasing need for data storage and retrieval. With lakhs of images added to the image database, not many images are annotated with proper description. So many relevant images go unmatched. The most widely accepted content-based image retrieval techniques cannot address the problems with all images, which are highly specialized .Our approach Histogram based Image Retrieval using Texture Feature retrieves the relevant images based on the texture property. We also provide an interface where the user can give a query image as an input. The texture feature is automatically extracted from the query image and is compared to the images in the database retrieving the matching images.

Author [α] : Assistant Professor in the department of CSE, Pydah College of engineering Kakinada, A.P, INDIA.
E-mail : neelima.sadineni@gmail.com
Author [σ] :Assistant Professor in Nimra Institute of Engineering & Technology, Ongole, Andhra Pradesh,India.
Author [ρ] : Assistant Professor in BVSR Engineering college, Chimakurthy, Ongole, Andhra Pradesh,India Anna University,Chennai
Author [ω] : Assistant Professor in Nimra Institute of Engineering & Technology,Ongole, Andhra Pradesh,India.
Author [¥] : Assistant Professor in Nalanda Institute of Engineering & Technology,Sathenapalli,Guntur, Andhra Pradesh,India.

II. Background & Previous Work

The goal of Content-Based Image Retrieval (CBIR) systems is to operate on collections of images and, in response to visual queries, extract relevant image. The application potential of CBIR for fast and effective image retrieval is enormous, expanding the use of computer technology to a management tool.

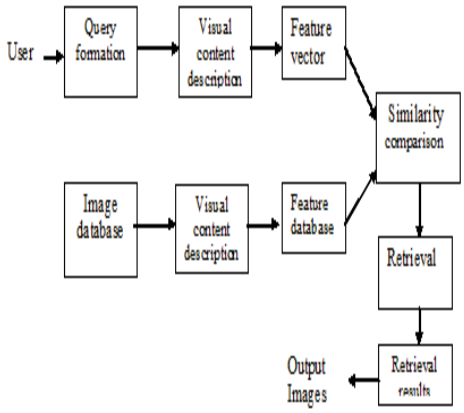


Fig.2 : Procedure for content based image retrieval system

CBIR operates on the principle of retrieving stored images from a collection by comparing features automatically extracted from the images themselves. The commonest features used are mathematical measures of color, texture or shape. A typical system allows users to formulate queries by submitting an example of the type of image being sought, though some offer alternatives such as selection from a palette or sketch input. The system then identifies those stored images whose feature values match those of the query most closely, and displays thumbnails of these images on the screen.

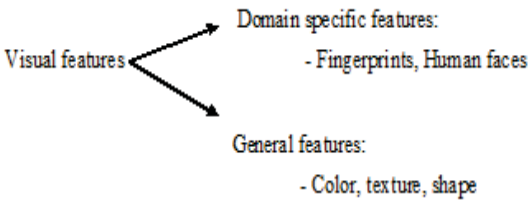


Fig.3 : Approaches of CBIR

Some of the most commonly used types of features used for image retrieval are as follows:

a) Colour Retrieval

Several methods for retrieving images on the basis of color similarity have been described in the literature, but most are variations on the same basic idea. Each image added to the collection is analyzed to compute a color histogram, which shows the proportion of pixels of each color within the image. The color histogram for each image is then stores in then stored in the database. At each time, the user can either specify the desired proportion of each color (&75% olive green and 25% red, for example), or submit an example image from which a color histogram is calculated. Either way, the matching process then retrieves those, which a color histogram is calculated. Either way, the matching process then retrieves those images whose color histograms match those of the query most closely.

b) Texture Retrieval

The ability to match on texture similarity can often be useful in distinguishing between areas of images with similar color (such as blue sky and sea or green leaves and grass). A variety of techniques has been used for measuring texture similarity; the best established rely on comparing values of what are known as second-order statistics calculated from query and stored images. Essentially, these calculate the relative brightness of selected pairs of pixels from each image. From these it is possible to calculate measures of image texture such as the degree of contrast, coarseness, directionality and regularity or periodically, directionality and randomness.

Texture queries can be formulated in a similar manner to color queries, by selecting examples of desires texture a palette, or by supplying an example query image. The system then retrieves images with texture measures most similar in value to the query.

c) Shape Retrieval

Two major steps are involves in shape feature extraction. They are object segmentation and shape representation.

Object segmentation: Segmentation is very important to Image Retrieval. Both the shape feature and the layout feature depend on good segmentation allow fast and efficient searching for information of a user's need.

d) Shape Representation

In image retrieval, depending on the applications, some requires the shape representation to be invariant to translation, rotation, and scaling. In general, the shape representations can be divided into two categories, boundary-based and region-based. The former uses only the outer boundary of the shape while the latter uses the entire shape region.

III. Concept of Texture Feature

Texture is one of the crucial primitives in human vision and texture features have been used to identify contents of images. Texture refers to the visual patterns that have properties of homogeneity that do not result from the presence of only a single color or intensity. Texture contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment. One crucial distinction between color and texture features is that color is a

point, or pixel, property, whereas texture is a local-neighbourhood property. As a result, it does not make sense to discuss the texture content at pixel level without considering the neighbourhood.

The texture is a property inherent to the surface. Various parameters or textural characteristics describe it. They are:

- The Granularity which can be rough or fine
- The Evenness which can be more or less good
- The Linearity
- The directivity
- The repetitiveness
- The contrast
- The order
- The connectivity

The other characteristics like color, size, and shape also must be considered. The Methodologies used for analysis of the texture are as follows

a) Texture Spectrum Method

The basic concept of texture spectrum method was introduced by H1 and Wang. The texture can be extracted from the neighborhood of 3 X 3 window which constitute the smallest unit called 'texture unit'. The neighborhood of 3 X 3 consists of nine elements respectively as V={ V1 , V2 , V3 , V4 , V0 ,V5 , V6 ,V7 , V8 } where V0 is the central pixel value and V1….V8 are the values of neighboring pixels within the window. The corresponding texture unit for this window is then a set containing eight elements surrounding the central pixel, represented as:

$$TU = \{ E1 , E2 , E3 , E4 , E0 , E5 , E6 , E7 , E8 \}$$

Where Ei is defined as: $Ei = \begin{cases} 0 & \text{if } Vi < V0 \\ 1 & \text{if } Vi = V0 \\ 2 & \text{if } Vi > V0 \end{cases}$

And the element E1 occupies the corresponding V1 pixel. Since each of the eight element of the texture units has any one of three values (0, 1, or 2)

$$NTU = \sum Ei * 3 (i - 1) \qquad [For\ i=1\ to\ 8]$$

Where NTU is the texture unit value. The occurrence distribution of texture unit is called the texture spectrum (TS). Each unit represents the local texture information of 3X3 pixels, and hence statistics of all the texture units in an image represent the complete texture aspect of entire image.

b) Cross Diagonal Texture Spectrum

AL-Jan obi (2001) has proposed a cross-diagonal texture matrix technique. In this method the eight neighboring pixels of 3 X 3 widows is broken up into two groups of four elements each at cross and diagonal positions. These groups are named as Cross Texture Unit (CTU) and Diagonal Texture Unit (DTU) respectively. Each of the four elements of these units is assigned a value (0, 1, and 2) depending on the gray level difference of the corresponding pixel with that of the central pixel of 3X3 window. These texture units have values from 0 to 80 (34, i.e 81 possible values).

Cross Texture Unit (CTU) and Diagonal Texture Unit (DTU) can be defined as:

$$NCTU = \sum Eci * 3 (i - 1) \qquad [For\ i=1\ to\ 4]$$

$$NDTU = \sum Edi * 3 (i - 1) \qquad [For\ i=1\ to\ 4]$$

Where NCTU and NDTU are the cross texture and diagonal texture unit values respectively; Eci and Edi are the ith elements of texture unit

The texture unit (CTU orDTU) value can range between :(0-240)

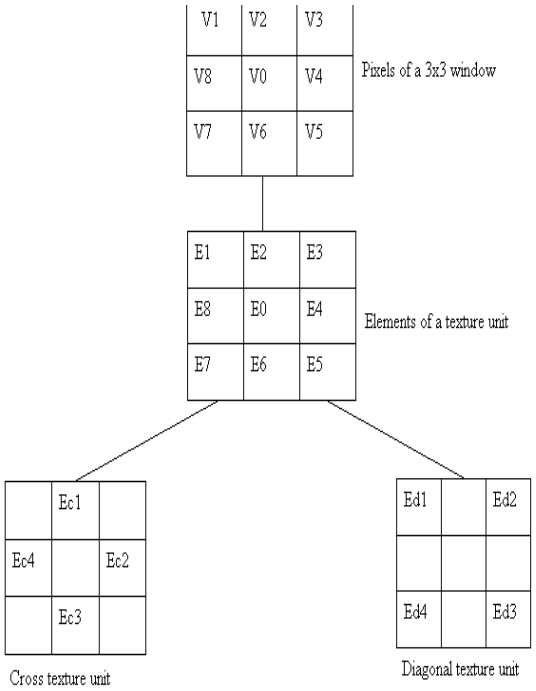


Fig.4 : Formation of cross diagonal texture units

c) Modified Texture Spectrum

In the proposed method, Nctu and Ndtu values have been evaluated which range from 0 to 80. For each type of texture unit, there can be four possible ways of ordering, which give four different values of CTU and DTU.

$$NTU = NCTU * NDTU$$

$$Ntu = Nctu + Ndtu$$

$$Ntu = Nctu - Ndtu$$

Where Nctu and Ndtu are the ordering ways for evaluation of Nctu and Ndtu.

After obtaining the CDTM values of 3*3 windows through entire image the occurrence frequency of each CDTM values are recorded. For the texture units having same CDTM values, two different procedures have been carried out to replace the pixel values of these units. The texture unit value can range between :(0-480).

d) Texture Spectrum with Thershold

The texture spectrum method with threshold is intended to make difference between the values of neighborhood matrix which are very close to the cental pixel value and those the rest.In this method the texture unit matix is represented as:

$$TU = \{ E1 , E2 , E3 , E4 , E0 , E5 , E6 , E7 , E8 \}$$

Where Ei is defined as: $Ei = 0$ if $Vi <= (V0 + t)$; 1 if $Vi > (V0 + t)$

Where t is the threshold value.

$$NTU = \sum Ei * 2 (i - 1) \quad [For\ i=1\ to\ 8]$$

The texture unit value can range between (0-254).

e) Reduced Texture Unit

In this method the range of texture unit values are (0,1).As the range is decreased the memory required to compute texture unit value also reduces.In this method

$$TU = \{ E1 , E2 , E3 , E4 , E0 , E5 , E6 , E7 , E8 \}$$

Where Ei is defined as: $Ei = 0$ if $Vi <= V0$; 1 if $Vi > V0$

Where t is the threshold value.

$$N\ RTU = \sum Ei * 2 (i - 1) \quad [For\ i=1\ to\ 8]$$

The texture unit value can range between (0-254).

f) Splitting Texture Unit Matrix into Rows and Columns

In this approach the texture unit matrix is split into 3 separate rows/columns. Texture unit value is calculated separately for each row/column. Later all the 3 texture unit values are added to get a single texture unit value. By doing this the texture unit value can be limited to 42.Thus memory and computation time can be saved.

E_1	E_2	E_3
E_8	E_0	E_4
E_7	E_6	E_5

Splitting into columns:

E_{11}	E_{21}	E_{31}
E_{12}	E_{22}	E_{32}
E_{13}	E_{23}	E_{33}

Here [E11- E13] are the first column values of the texture unit matrix denoted as TU1. Similarly [E21- E23] and [E31- E33] denote second (TU2) and third (TU3) columns of texture unit matrix respectively.

The texure unit value is calculated separately for each texture unit matrix (j) as:

$$N\ TUj = \sum Eji * 2 (i - 1) \quad [For\ i=1\ to\ 3]$$

The final texture unit value is evaluated as:

$$N\ TU = \sum N\ TUj \quad [For\ j=1\ to\ 3]$$

The texture unit value can range between (0-42).

Splitting into rows:

E_{11}	E_{12}	E_{13}
E_{21}	E_{22}	E_{23}
E_{31}	E_{32}	E_{33}

Here [E11- E13] are the first row values of the texture unit matrix denoted as TU1. Similarly [E21- E23] and [E31- E33] denote second (TU2) and third (TU3) rows of texture unit matrix respectively.

The texure unit value is calculated separately for each texture unit matrix (j) as:

$$N\ TUj = \sum Eji * 2 (i - 1) \quad [For\ i=1\ to\ 3]$$

The final texture unit value is evaluated as:

$$N\ TU = \sum N\ TUj \quad [For\ j=1\ to\ 3]$$

The texture unit value can range between (0-42).

IV. Histogram Intersection Approach

To overcome the disadvantages of Euclidean distance we taken histogram intersection measure. The histogram intersection was investigated for color image retrieval by swain and Ballard. Their objective was to find known objects within images using color histograms. When the object (q) size is less than the image (t) size, and the histograms are not normalized, then $|hq| <= |ht|$. The intersection of histograms hq and ht is given by:

$$d_{q,t} = 1 - \frac{\sum_{m=0}^{M-1} \min (h_q[m], h_t[m])}{|h_q|}$$

Where $|h| = \Sigma\ h[m]$ [for m=0 to M-1].The above equation is not a valid distance metric since it is not symmetric hq,t not equal to dt,q. However that equation can be modified to produce a true distance metric by making it symmetric in hq and ht as follows:

$$d^{1}_{q,t} = 1 - \frac{\sum_{m=0}^{M-1} \min(h_q[m], h_t[m])}{\min(|h_q|, |h_t|)}$$

Alternatively when the histograms are normalized such that $|hq| = |ht|$, both equations are true distance metrics.When $|hq| = |ht|$ that $D1(q,t) = dq,t$ and the Histogram Intersection is given by

$$D1_{(q,t)} = \sum |h_q[m] - h_t[m]| \quad [\text{For } m=0 \text{ to } M-1]$$

V. Design & Analysis of The Experiment

Class Diagram models class structure and contents using design elements such as classes, packages and objects as shown in Fig.5. It also displays relationships such as containment, inheritance, associations and others.

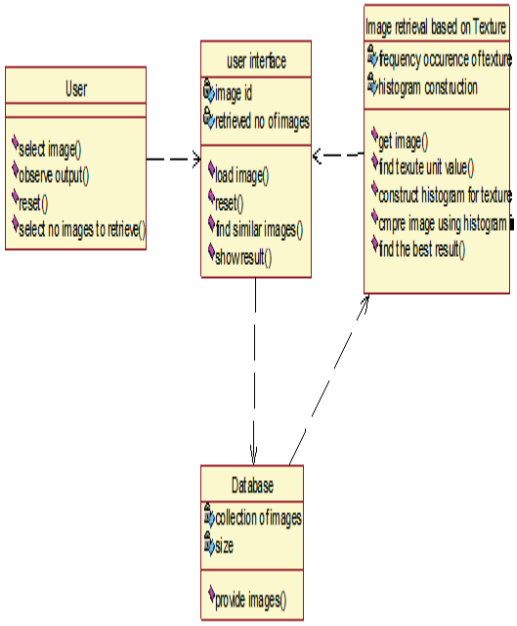


Fig.5 : Interpretational Class Diagram for User interaction

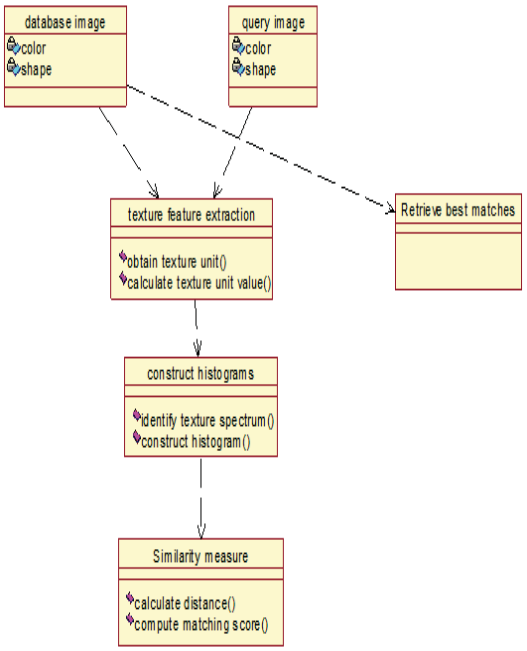


Fig.6 : Interpretational Class Diagram for Image Retrieval

Sequence Diagram displays the time sequence of the objects participating in the interaction as shown in Fig.6. This consists of the vertical dimension (time) and horizontal dimension (different objects).

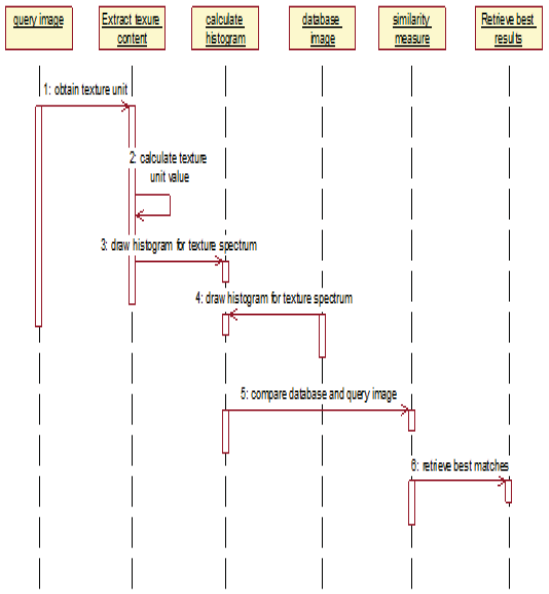


Fig.7 : Interpretational Sequence Diagram of Database search

VI. Results

The below are the results obtained from the experiment

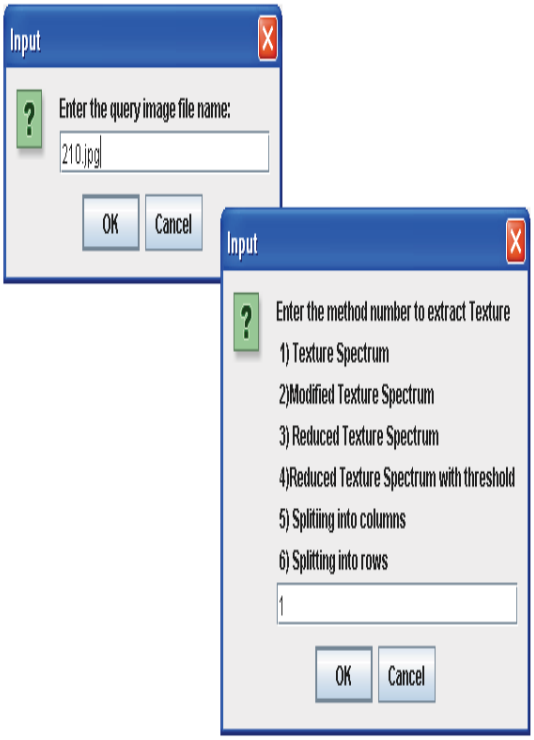


Fig.8 : Texture Spectrum Method

Fig.9 : Modified Texture Spectrum Method

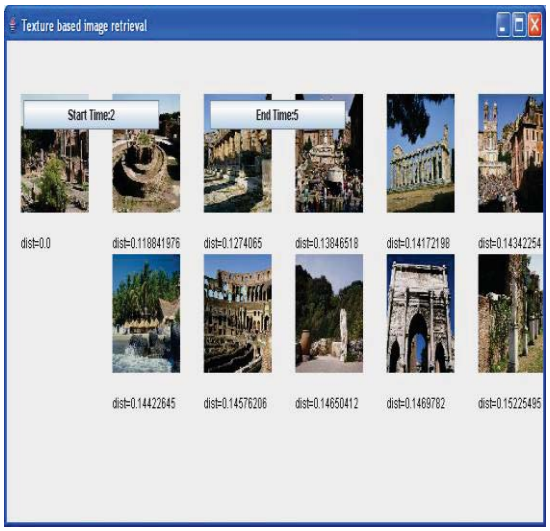


Fig.10 : Reduced Texture Spectrum Method

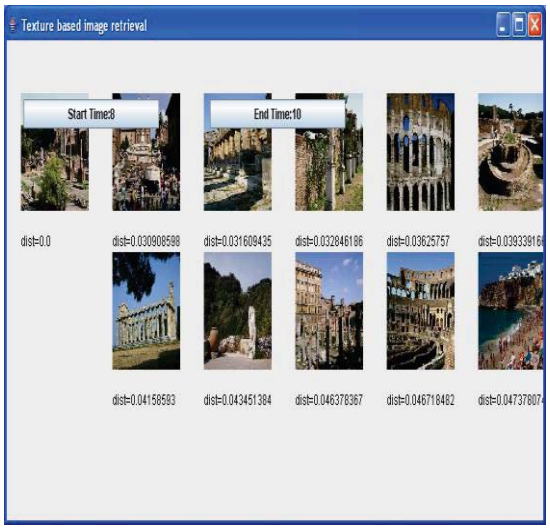


Fig.11: Reduced Texture Spectrum with Threshold

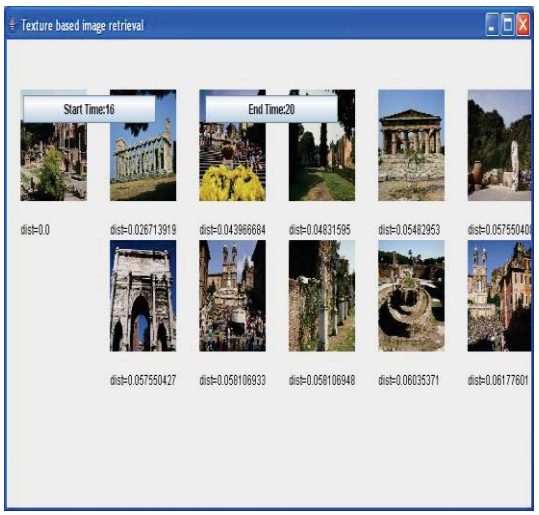


Fig.12 : Splitting into Rows

VII. Conclusion & Future Enhancement

In this work we experimented with the ideas of Histogram based Image Retrieval using Texture Feature system with different methods of extracting texture feature. We incorporated the Histogram Intersection measure method to compare the query image with database images. A measure of the over-all similarity between images, defined by our approach, incorporates all local properties of the texture histograms of the images. We proved that our approach is well suited to retrieve best possible results. There are several improvements that can be taken as future work for this project. Our system considers only the texture feature of the image. Consideration of other features like shape, location can help for a better retrieval of images. The database of images can be of even more images.

References Références Referencias

1. F. Korn, N. Sidiropoulos, C. Faloustos, E. Siegel, and Z. Protopapas. Fast and e.ective retrieval of medical tumor shapes. IEEE Transactions on Knowledge and Data Engineering, 10(6):889–904, 1998.
2. T. Lehmann, B. Wein, J. Dahmen, J. Bredno,F. Vogelsang, and M. Kohnen. Content-Based Image Retrieval in Medical Applications: A Novel Multi-Step Approach. In International Society for Optical Engineering (SPIE), volume 3972(32), pages 312–320, Feb. 2000.
3. W. Y. Ma and B. S. Manjunath. Netra: a toolbox for navigating large image databases. In International Conference on Image Processing (ICIP), pages 568–571, Oct. 1997.
4. R. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classi.cation. IEEE Transactions Systems on Man and Cybernetics, 3(6):610 621, 1973.
5. T. S. Huang, S. Mehrotra, and K. Ramchandran. Multimedia Analysis and Retrieval System (MARS) project. In Proceedings of 33rd Annual Clinic on Library Application of Data processing - Digital Image Access and Retrieval, pages 100–117, Urbana/Champaign, Illinois, 1996.
6. K. Jain and R. Dubes. Algorithms for clustering data. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.
7. Kak and C. Pavlopoulou. Content-Based Image Retrieval from Large Medical Databases. In 3D Data Processing, Visualization, Transmission, Padova, Italy, June 2002.
8. W. Chu, C. Hsu, A. Cardenas, and R. T aira. Knowledge-based image retrieval with spatial and temporal constructs. IEEE Transactions on Knowledge and Data Engineering, 10(6):872–888, 1998.
9. Clausi and M. Jernigan. Designing Gabor filters for optimal texture separability. Pattern Recognition, 33:1835–1849, Jan. 2000.
10. Comaniciu, P. Meer, D. Foran, and E. Medl. Bimodal system for interactive indexing and retrieval of pathology images. In Workshop on Applications of Computer Vision, pages 76–81, Princeton, NJ, Oct.1998.
11. J. Bach, C. Fuller, A. Gupta, A. Hampapur,B. Horowitz, R. Humphrey, and R. Jain. Virage image search engine: An open framework for imagemanagement. In SPIE Storage and Retrieval for Imageand Video Databases, volume 2670, pages 76–87, SanDiego/La Jolla, CA, Jan. 1996.
12. Bovik, M. Clark, and W. Geisler. Multichannel texture analysis using localized spatial .lters. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12:55–73, 1990.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 5 Version 1.0 March 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Content Based Data Retrieval on KNN- Classification and Cluster Analysis for Data Mining

By Aws Saad Shawkat & H K Sawant
Bharati Vidyapeeth Deemed University Pune

Abstract - Data mining is sorting through data to identify patterns and establish relationships.Data mining parameters include: Regression -In statistics, regression analysisincludes any techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. Sequence or path analysis -looking for patterns where one event leads to another later event. Classification -looking for new patterns. Clustering -finding and visually documenting groups. Decision Trees – Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal.

GJCST Classification: I.5.3

Strictly as per the compliance and regulations of:

Content Based Data Retrieval on KNN-Classification and Cluster Analysis for Data Mining

Aws Saad Shawkat [α] & H K Sawant [α]

Abstract - Data mining is sorting through data to identify patterns and establish relationships.Data mining parameters include: Regression - In statistics, regression analysis includes any techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. Sequence or path analysis - looking for patterns where one event leads to another later event. Classification - looking for new patterns. Clustering - finding and visually documenting groups. Decision Trees – Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal.

I. Introduction

Data mining is an iterative process that typically involves the following phases:

a) *Problem definition* : A data mining project starts with the understanding of the business problem. Data mining experts, business experts, and domain experts work closely together to define the project objectives and the requirements from a business perspective. The project objective is then translated into a data mining problem definition. In the problem definition phase, data mining tools are not yet required.

b) *Data exploration* : Domain experts understand the meaning of the metadata. They collect, describe, and explore the data. They also identify quality problems of the data. A frequent exchange with the data mining experts and the business experts from the problem definition phase is vital.

In the data exploration phase, traditional data analysis tools, for example, statistics, are used to explore the data.

c) *Data preparation* : Domain experts build the data model for the modeling process. They collect, cleanse, and format the data because some of the mining functions accept data only in a certain format. They also create new derived attributes, for example, an average value. In the data preparation phase, data is tweaked multiple times in no prescribed order. Preparing the data for the modeling tool by selecting tables, records, and attributes, are typical tasks in this phase. The meaning of the data is not changed.

d) *Modeling* : Data mining experts select and apply various mining functions because you can use different mining functions for the same type of data mining problem. Some of the mining functions require specific data types. The data mining experts must assess each model. In the modeling phase, a frequent exchange with the domain experts from the data preparation phase is required. The modeling phase and the evaluation phase are coupled. They can be repeated several times to change parameters until optimal values are achieved. When the final modeling phase is completed, a model of high quality has been built.

e) *Evaluation* : Data mining experts evaluate the model. If the model does not satisfy their expectations, they go back to the modeling phase and rebuild the model by changing its parameters until optimal values are achieved. When they are finally satisfied with the model, they can extract business explanations and evaluate the following questions: Does the model achieve the business objective? Have all business issues been considered? At the end of the evaluation phase, the data mining experts decide how to use the data mining results.

f) *Deployment* : Data mining experts use the mining results by exporting the results into database tables or into other applications, for example, spreadsheets.

II. Data Mining Types

a) Predictive Data Mining: Predictive data mining involves creation of model system based on and described by a given set of data.

b) Descriptive Data Mining: Descriptive data mining on the other hand produces new and unique information inferred from the available set of data.

Raw Data: Raw data is a term for data collected on source which has not been subjected to processing or any other manipulation. (Primary data), it is also known as primary data. It is a relative term (see data). Raw data can be input to a computer program or used in manual analysis procedures such as gathering statistics from a survey. It can refer to the binary data on electronic storage devices such as hard disk drives (also referred to as low-level data).

Author [α] *: Department of Information Technology Bharati Vidyapeeth Deemed University College of Engineering, Pune-46.*
E-mail : awsit85@gmail.com

a) Normalization of Raw Data

Some data-mining methods, typically those that are based on distance computation between points in an n-dimensional space, may need normalized data for best results.

Here are three simple and effective normalization techniques:

b) Decimal Scaling

Decimal scaling moves the decimal point but still preserves most of the original digit value.

$$VI' = VI/10^K$$

c) Min-Max Normalization

Suppose that the data for a feature v are in a range between 150 and 250. Then, the previous method of normalization will give all normalized data between .15 and .25; but it will accumulate the values on a small subinterval of the entire range. To obtain better distribution of values on a whole, normalized interval, e.g., [0, 1], we can use the min-max formula

$$VI' = (VI - Min(VI))/(Max(VI) - Min(VI))$$

d) Standard Deviation Normalization

Normalization by standard deviation often works well with distance measures, but transforms the data into a form unrecognizable from the original data.

$$VI' = (VI - Mean(V))/Std(V)$$

Types of Data

Categorical Data: Categorical data (or variable) consists of names representing categories. For example, the gender (categories of male & female) of the people where you work or go to school; or the make of cars in the parking lot (categories of Ford, GM, Toyota, Mazda, KIA, etc) is categorical data.

Numerical Data: Numerical data (or variable) consists of numbers that represent counts or measurements. For example, the number of males & females where you work or go to school; or the number of the make of cars Ford, GM, Toyota, Mazda, KIA, etc is numerical data.

Dummy Variable: A dummy variable is a numerical variable used in regression analysis to represent subgroups of the sample in your study.

Discrete Variable: Discrete Variable are also called Qualitative Variable. It is nominal or ordinal.

Continuous Variable: Continuous variable are measured using interval scale or ratio scale.

III. Data Reduction

Means reducing the number of cases or variables in a data matrix. The basic operations in a data-reduction process are delete column, delete a row, and reduce the number of values in a column. These operations attempt to preserve the character of the original data by deleting data that are nonessential. There are other operations that reduce dimensions, but the new data are unrecognizable when compared to the original data set, and these operations are mentioned here just briefly because they are highly application-dependent.

a) Entropy

A method for unsupervised feature selection or ranking based on entropy measure is a relatively simple technique; but with a large number of features its complexity increases significantly.

The similarity measure between two samples can be defined as

$$\alpha = -(\ln 0.5)/D$$

$$S_{ij} = e^{-\alpha D_{ij}}$$

Where Dij is the distance between the two samples xi and xj and α is a parameter mathematically expressed as

D is the average distance among samples in the data set. Hence, α is determined by the data. But, in a successfully implemented practical application, it was used a constant value of α = 0.5. Normalized Euclidean distance measure is used to calculate the distance Dij between two samples xi and xj:

$$D_{ij} = \left[\sum_{k=1}^{n}\left((x_{ik} - x_{jk})/(\max(k) - \min(k))\right)^2\right]^{1/2}$$

- where n is the number of dimensions and max(k) and min(k) are maximum and minimum values used for normalization of the k-th dimension.
- All features are not numeric. The similarity for nominal variables is measured directly using Hamming distance.

$$S_{ij} = \left(\sum_{k=1}^{n}|x_{ik} - x_{jk}|\right)/n$$

where

The total number of variables is equal to n. For mixed data, we can discretize numeric values (Binning) and transform numeric features into nominal features before we apply this similarity measure.

If the two measures are close, then the reduced set of features will satisfactorily approximate the original set. For a data set of N samples, the entropy measure is

$$E = -\sum_{i=1}^{N-1}\sum_{j=1}^{N}\left(S_{ij}\times\log S_{ij} + (1 - S_{ij})\times\log(1 - S_{ij})\right)$$

where Sij is the similarity between samples xi and xj. This measure is computed in each of the iterations as a basis for deciding the ranking of features. We rank features by gradually removing the least important feature in maintaining the order in the

configurations of data. The steps of the algorithm are base on sequential backward ranking, and they have been successfully tested on several real-world applications.

b) Linear Regreesion

In statistics, linear regression refers to any approach to modeling the relationship between one or more variables denoted y and one or more variables denoted X, such that the model depends linearly on the unknown parameters to be estimated from the data.

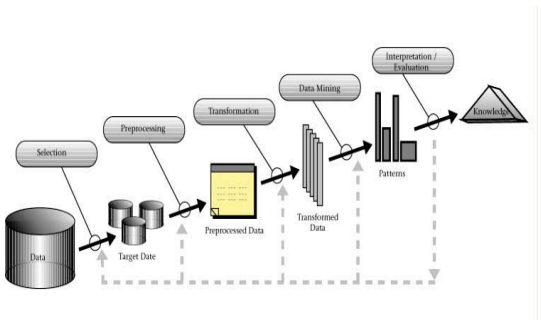
Linear regression has many practical uses. Most applications of linear regression fall into one of the following two broad categories:

- If the goal is prediction, or forecasting, linear regression can be used to fit a predictive model to an observed data set of *y* and *X* values. After developing such a model, if an additional value of *X* is then given without its accompanying value of *y*, the fitted model can be used to make a prediction of the value of *y*.
- Given a variable *y* and a number of variables X_1, ..., X_p that may be related to *y*, then linear regression analysis can be applied to quantify the strength of the relationship between *y* and the X_j, to assess which X_j may have no relationship with *y* at all, and to identify which subsets of the X_j contain redundant information about *y*, thus once one of them is known, the others are no longer informative.

IV. Implementation

The core task of Data Mining Model is the application of the appropriate mining function to your data to build mining models that answer your business questions. Administrative tasks such as retrieving progress information or interpreting error messages support this task.

Data Mining Process



a) State the Problem

A data mining project starts with the understanding of the problem. Data mining experts and domain experts work closely together to define the project objectives and the requirements from a business perspective. The project objective is then translated into a data mining problem definition.

b) Data Normalization

The use of the data transformation in my project is to make the data symmetric. In practice a suitable data transformation can be selected by examining the effect of the transformation.. So for the medical data set min-max transformation is often used.

$$VI' = (VI-Min(VI))/(Max(VI)-Min(VI))$$

c) Missing Values Adjustment

The Missing value technique used in these type of project is to take the mean of that feature but the data set which I have choose for the project have no missing values.

d) Outlier Analysis

The technique used by data set to remove the outlier values is the Deviation based technique in which the human can easily distinguish unusual samples from a set of other similar samples.

After examining each and every data cluster, we obtain data set which contains no outlier.

e) Data Reduction

The term data reduction in the context o data mining is usually applied to projects where the goal is to aggregate the information contained in large data sets into manageable(smaller) information nuggets. Data reduction method can include simple tabulation ,aggregation (computing descriptive statistics) or more sophisticated technique like principle component analysis.

Since the data which I have used in the project is not so huge therefore there is no need of applying the data reduction because it could lead to the loss of information from the data.

f) Model Estimation

A model can be defined as a number of examples or a mathematical relationship. Data mining experts select and apply various mining functions because we can use different mining functions for the same type of data mining problem. Some of the mining functions require specific data types.

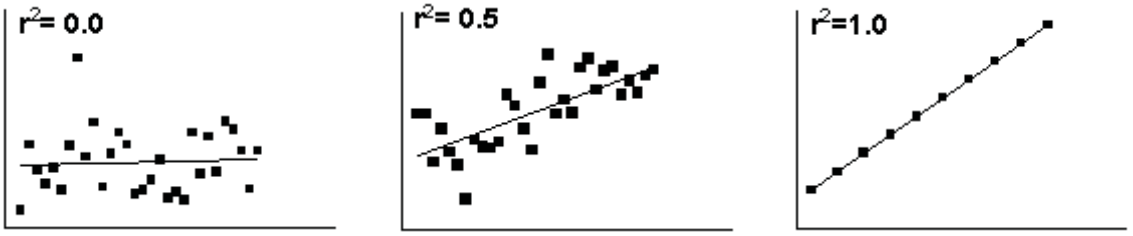
g) Linear Regression

Regression: The purpose of this model function is to map a data item to a real-valued prediction variable.

The goal of regression is to build a concise model of the distribution of the dependent attribute in terms of the predictor attributes. The resulting model is used to assign values to a database of testing records, where the values of the predictor attributes are known but the dependent attribute is to be determined.

The value r^2 is a fraction between 0.0 and 1.0, and has no units. An r^2 value of 0.0 means that knowing X does not help you predict Y. There is no linear relationship between X and Y, and the best-fit line is a horizontal line going through the mean of all Y values.

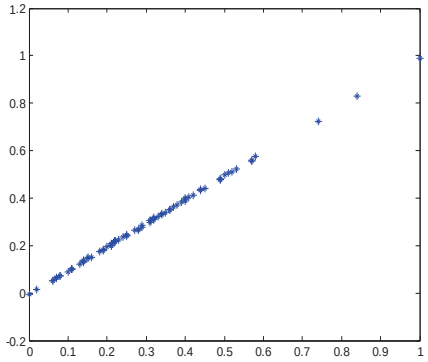
When r^2 equals 1.0, all points lie exactly on a straight line with no scatter. Knowing X lets you predict Y perfectly.



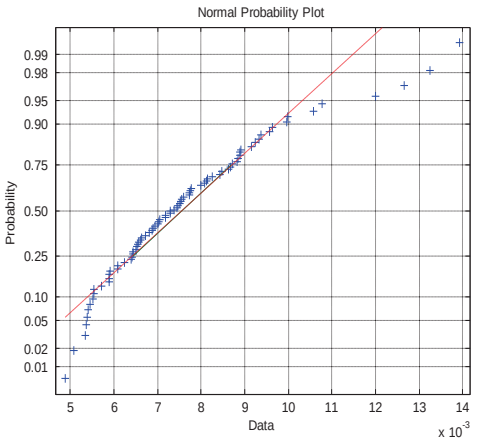
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.995[a]	.990	.989	.01950

Since the error is very small so the result which we get after applying is very close to the final result. The graph between observed and fitted value is shown in figure



The normal probability plot is a special case of the probability plot. We cover the normal probability plot separately due to its importance in many applications. The normal probability plot is formed by:
Vertical axis: Ordered response values
Horizontal axis: Normal order statistic medians
The normal probability plot is shown in the figure



h) Cluster Analysis: Cluster analysis or clustering is the assignment of a set of observations into subsets (called clusters) so that observations in the same cluster are similar in some sense. Clustering is a method of unsupervised learning, and a common technique for statistical data analysis used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics Applying hierarchical clustering algorithm.

i) Hierarchical Clustering

It begins with as many clusters as objects. Clusters are successively merged until only one cluster remains.

- Representation of all pair-wise distances
- Parameters: none (distance measure)
- Results
- One large cluster
- Hierarchical tree (dendrogram)
- Deterministic
- Agglomerative: This is a "bottom up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.
- Divisive: This is a "top down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy

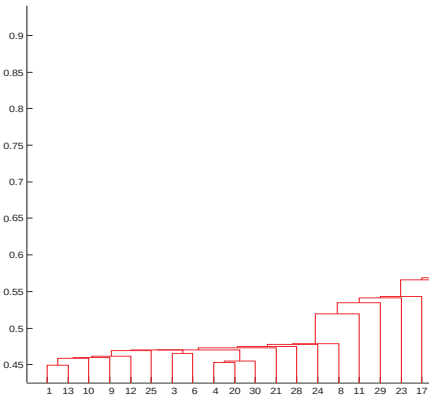
The K-means partitional-clustering algorithm is the simplest and most commonly used algorithm employing a square-error criterion.

It starts with a random, initial partition and keeps reassigning the samples to clusters, based on the similarity between samples and clusters, until a convergence criterion is met.

- Partition data into K clusters
- Parameter: Number of clusters (K) must be chosen

- Randomized initialization:
- Different clusters each time
- Non-deterministic

Here the k-mean is applied to calculate the final cluster centers among samples.



V. Conclusion

The model in which every decision is based on the comparison of two numbers within constant time is called simply a decision tree model. It was introduced to establish computational complexity of sorting and searching,__advantages of applying is Easy to understand, Map nicely to a set of business rules, Applied to real problems, Make no prior assumptions about the data, Able to process both numerical and categorical data.

Data mining techniques are used in a many research areas, including mathematics, cybernetics, genetics and marketing. Web mining, a type of data mining used in customer relationship management (CRM), takes advantage of the huge amount of information gathered by a Web site to look for patterns in user behavior.

References Références Referencias

1. Jinshu, Zhang Bofeng, Xu Xin, "Advances in Machine Learning Based Text Categorization", *Journal of Software*, Vol.17, No.9, 2006, pp.1848 1859
2. Ma Jinna, "Study on Categorization Algorithm of Chinese Text", *Dissertation of Master's Degree*, University of Shanghai for Science and Technology, 2006
3. Wang Jianhui, Wang Hongwei, Shen Zhan, Hu Yunfa, "A Simple and Efficient Algorithm to Classify a Large Scale of Texts", *Journal of Computer Research and Development*, Vol.42, No.1, 2005, pp.85 93
4. Li Ying, Zhang Xiaohui, Wang Huayong, Chang Guiran, "Vector-Combination-Applied KNN Method for Chinese Text Categorization", *Mini- Micro Systems*, Vol.25, No.6, 2004, pp.993 996
5. Wang Yi, Bai Shi, Wang Zhang'ou, "A Fast KNN Algorithm Applied to Web Text Categorization", *Journal of The China Society for Scientific and Technical Information*, Vol.26, No.1, 2007, pp.60 64
6. Fabrizio Sebastiani, "Machine learning in automated text categorization", *ACM Computer Survey*,Vol.34, No.1, 2002, pp. 1-47
7. Lu Yuchang, Lu Mingyu, Li Fan, "Analysis and construction of word weighing function in vsm", *Journal of Computer Research and Development*, Vol.39, No.10, 2002, pp.1205 1210
8. Belur V, Dasarathy, "Nearest Neighbor (NN) Norms NN Pattern Classification Techniques", *Mc Graw-Hill Computer Science Series, IEEE Computer Society Press*, Las Alamitos, California, 1991,pp.217-224
9. Yang Lihua , Dai Qi, Guo Yanjun, "Study on KNN Text Categorization Algorithm", *Micro Computer Information*, No.21, 2006, pp.269 271
10. Wang Yu, Wang Zhengguo, „A fast knn algorithm for text categorization", *Proceedings of the Sixth International Conference on Machine Learning and Cybernetics*, Hong Kong, 19-22 August 2007, pp.3436-3441
11. Yang Y, Pedersen J O, "A comparative study on feature selection in text categorization", *ICNL*, 1997,pp.412-420
12. Xinhao Wang, Dingsheng Luo, Xihong Wu, Huisheng Chi, "Improving Chinese Text Categorization by Outlier Learning", *Proceeding ofNLP-KE'05* pp. 602-607
13. Jin Yang, Zuo Wanli, "A Clustering Algorithm Using Dynamic Nearest Neighbors Selection Model", *Chinese Journal of Computers*, Vol.30, No.5, 2005, pp.759 762

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 5 Version 1.0 March 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

An Efficient Approach of Removing the High Density Salt & Pepper Noise Using Stationary Wavelet Transform

By N.Naveen Kumar & Dr.S.Ramakrishna
S.V. University, Tirupati Andhra Pradesh, India

Abstract - Images are often corrupted by impulse noise, also known as salt and pepper noise. Salt and pepper noise can corrupt the images where the corrupted pixel takes either maximum or minimum gray level. Amongst these standard median filter has been established as reliable - method to remove the salt and pepper noise without harming the edge details. However, the major problem of standard Median Filter (MF) is that the filter is effective only at low noise densities. When the noise level is over 50% the edge details of the original image will not be preserved by standard median filter. Adaptive Median Filter (AMF) performs well at low noise densities. In our proposed method, first we apply the Stationary Wavelet Transform (SWT) for noise added image. It will separate into four bands like LL, LH, HL and HH. Further, we calculate the window size 3x3 for LL band image by Reading the pixels from the window, computing the minimum, maximum and median values from inside the window. Then we find out the noise and noise free pixels inside the window by applying our algorithm which replaces the noise pixels. The higher bands are smoothing by soft thresholding method. Then all the coefficients are decomposed by inverse stationary wavelet transform. The performance of the proposed algorithm is tested for various levels of noise corruption and compared with standard filters namely standard median filter (SMF), weighted median filter (WMF). Our proposed method performs well in removing low to medium density impulse noise with detail preservation up to a noise density of 70% and it gives better Peak Signal-to-Noise Ratio (PSNR) and Mean square error (MSE) values.

GJCST Classification: G.1.2

Strictly as per the compliance and regulations of:

An Efficient Approach of Removing the High Density Salt & Pepper Noise Using Stationary Wavelet Transform

N.Naveen Kumar[α] & Dr.S.Ramakrishna[σ]

Abstract - Images are often corrupted by impulse noise, also known as salt and pepper noise. Salt and pepper noise can corrupt the images where the corrupted pixel takes either maximum or minimum gray level. Amongst these standard median filter has been established as reliable - method to remove the salt and pepper noise without harming the edge details. However, the major problem of standard Median Filter (MF) is that the filter is effective only at low noise densities. When the noise level is over 50% the edge details of the original image will not be preserved by standard median filter. Adaptive Median Filter (AMF) performs well at low noise densities. In our proposed method, first we apply the Stationary Wavelet Transform (SWT) for noise added image. It will separate into four bands like LL, LH, HL and HH. Further, we calculate the window size 3x3 for LL band image by Reading the pixels from the window, computing the minimum, maximum and median values from inside the window. Then we find out the noise and noise free pixels inside the window by applying our algorithm which replaces the noise pixels. The higher bands are smoothing by soft thresholding method. Then all the coefficients are decomposed by inverse stationary wavelet transform. The performance of the proposed algorithm is tested for various levels of noise corruption and compared with standard filters namely standard median filter (SMF), weighted median filter (WMF). Our proposed method performs well in removing low to medium density impulse noise with detail preservation up to a noise density of 70% and it gives better Peak Signal-to-Noise Ratio (PSNR) and Mean square error (MSE) values.

I. Introduction

Impulse noise may often corrupt the images, which is known as salt and pepper noise..A standard signal processing requirement is to remove randomly occurring impulses without disturbing the edges. It is well known that linear filtering techniques fail when the noise is non-additive and are not effective in removing impulse noise. This lead researchers to make use of the nonlinear signal processing techniques. Based on two types of image models corrupted by impulse noise, two new algorithms for adaptive median filters are presented in Ref. [1]. these have variable window size for removal of impulses while preserving sharpness. The first one, called the ranked-order based adaptive median filter (RAMF), is based on a test for the presence of impulses in the center pixel itself followed by the test for the presence of residual impulses in the median filter output. The second one, called the impulse size based adaptive median filter (SAMF), is based on the detection of the size of the impulse noise.

A new impulse noise detection technique for switching median filters was described in Ref. [2], which is based on the minimum absolute value of four convolutions obtained using one-dimensional Laplacian operators. Extensive simulations show that the proposed filter provides better performance than many of the existing switching median filters with comparable computational complexity.

Srinivasan et al. [3], proposed a new decision-based algorithm for the restoration of images that are highly corrupted by impulse noise. They reported significantly better image quality than a standard median filter (SMF), adaptive median filters (AMF), threshold decomposition filter (TDF), cascade, and recursive nonlinear filters. Unlike other nonlinear filters, this method, removes only corrupted pixel by the median value or by its neighboring pixel value.

Previously, many linear and nonlinear filtering techniques have been described to remove impulse noise. However, these filters often bring along blurred and distorted image of details. A detail preserving filter for impulse noise removal was proposed by Dagao Duan et al. [4]. on the basis of the Soft-Switching Median (SWM) filter. Moreover, Eduardo Abreu [5] reported a new framework for removing impulse noise from images, in which the nature of the filtering operation is conditioned on a state variable defined as the output of a classifier that operates on the differences between the input pixel and the remaining rank-ordered pixels in a sliding window. As part of this framework, several algorithms are examined, each of which is applicable to fixed and random-valued impulse noise models. Also, Chenhen et al. [6] reported a novel nonlinear filter, called tri-state median (TSM) filter, for preserving image details while effectively suppressing impulse noise. The standard median (SM) filter and the center weighted median (CWM) filter into a noise detection framework to determine whether a pixel is corrupted, before applying filtering unconditionally.

Author[α] : Research Scholar, Department of Computer Science, S.V. University, Tirupati -517 502, Andhra Pradesh, India.
E-mail : naveennsvu@gmail.com
Author[σ] : Professor, Department of Computer Science, S.V. University, Tirupati -517 502, Andhra Pradesh, India.
E-mail : drsramakrishna@yahoo.com

To restore images corrupted by salt–pepper impulse noise, a new median-based filter such as progressive switching median (PSM) filter was presented in Ref. [7]. It was developed on the basis of the following two main points: 1) switching scheme an impulse detection algorithm is used before filtering, thus only a proportion of all the pixels will be filtered and, 2) progressive methods both the impulse detection and the noise filtering procedures are progressively applied through several iterations.

A generalized framework of median based switching schemes, called multi- state median (MSM) filter is presented in Ref. [8]. By using simple thresholding logic, the output of the MSM filter is adaptively switched among those of a group of center weighted median (CWM) filters that have different center weights. A novel switching-based median filter with incorporation of fuzzy-set concept, called the noise adaptive soft-switching median (NASM) filter [9], to achieve much improved filtering performance in terms of effectiveness in removing impulse noise while preserving.

signal details and robustness in combating noise density variations. Also, Luo et al [10] designed a new efficient algorithm for the removal of impulse noise from corrupted images while preserving image details. It was interpreted on the basis of the alpha-trimmed mean, which is a special case of the order-statistics filter.

II. Methodology

The proposed image denoising corrupted by salt and pepper noise is built on Stationary wavelet transform. The following section describe the Stationary Wavelet transform (SWT) and Proposed Algorithm used in the present work (Figure.1).

a) Stationary Wavelet Transform

The SWT provides efficient numerical solutions in the signal processing applications. It was independently developed by several researchers and under different names, e.g. the un-decimated wavelet transform, the invariant wavelet transform and the redundant wavelet transform. The key point is that it gives a better approximation than the discrete wavelet transform (DWT) since, it is redundant, linear and shift invariant. These properties provide the SWT to be realized using a recursive algorithm. Thus, the SWT is a very useful algorithm for analyzing a linear system.

A brief description of the SWT is presented here. It shows the computation of the SWT of a signal$x(k)$, whereW_{jk}, and V_{jk} are called the detail and the approximation coefficients of the SWT. The filters H_j and G_j are the standard lowpass and highpass wavelet filters, respectively. In the first step, the filters H_1 and G_1 are obtained by upsampling the filters using the previous step (i.e. $H_j - 1$ and$G_j - 1$).

Block Diagram

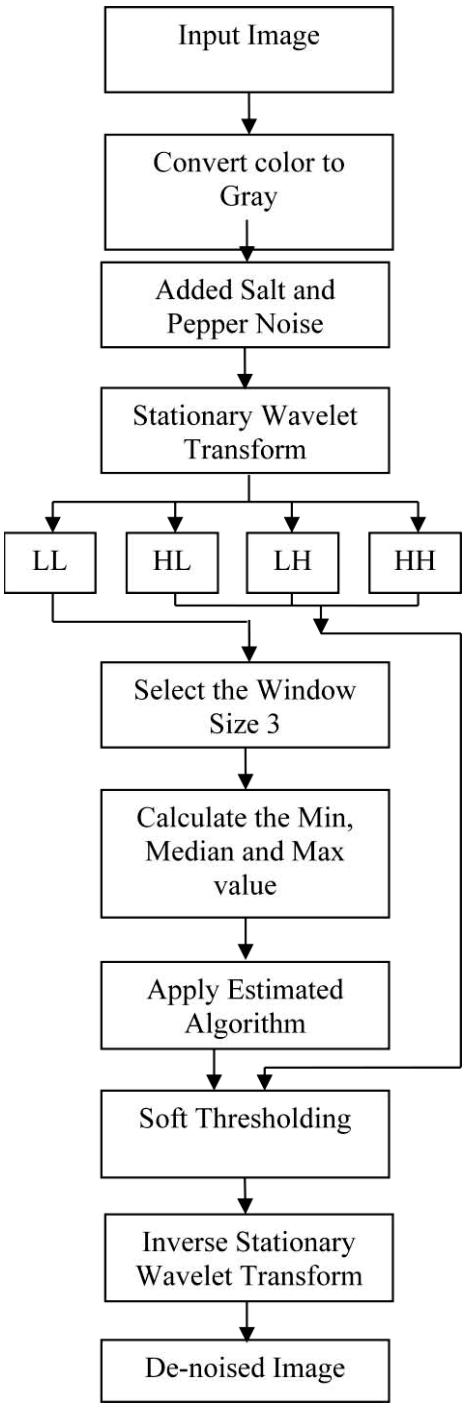


Fig 1: Block Diagram of Proposed Method

III. Proposed Method

Let X denote the noise corrupted image, apply the stationary wavelet transform of the noise image to get the LL, LH, HL and HH bands. In LL band for each pixel $X(i,j)$ denoted as X_{ij} a sliding or filtering window of size $2(L+1)X(2L+1)$ centered at X_{ij} is defined as shown in Table.1. The elements of this window are $S_{ij} = \{X_{i-u,j-v,} - L \leq u,v \leq L\}$.

X(i-1,j-1)	X(i-1,j)	X(i-1,j+1)
X(i,j-1)	X(i,j)	X(i,j+1)
X(i+1,j+1)	X(i+1,j)	X(i+1,j+1)

Table.1: 3 x 3 Filtering window with X (i,j) as center pixel

1. Set the minimum window size w=3;
2. Read the pixels from the sliding window and store it in(S).
3. Compute minimum, maximum and median value inside the window.
4. If the center pixel in the window X(i,j),is such that min<X(i,j)<max, then it is considered as uncorrupted pixel and retained. Otherwise go to step 5.
5. Select the pixels in the window such that min<Sij<max if number of pixels is less than 1 then increase the window size by 2 and go to step2 ,else go to step 6.
6. Difference of each pixel inside the window with the median value is calculated as x and applied to robust influence function.

$$f(x) = \frac{2x}{(2\sigma^2 + x^2)} \tag{1}$$

where σ is outlier rejection point which is given by

$$\sigma = \frac{\tau_s}{\sqrt{2}} \tag{2}$$

where $\tau_s$the maximum is expected outlier and is given by

$$\tau_s = \varsigma\sigma_N \tag{3}$$

where σ_N the local is estimate of the image standard deviation and ς is a smoothening factor. Here δ =0.3 is taken for medium smoothening.

7. Pixel is estimated using equation (4) and (5)

$$S1 = \sum_{l\in L} \frac{pixel(l) * f(x)}{x} \tag{4}$$

$$S2 = \sum_{l\in L} \frac{f(x)}{x} \tag{5}$$

For all higher bands (LH, HL and HH) the de-noising can be achieved by applying a thresholding operator to the wavelet coefficients in the transform domain followed by reconstruction of the signal to the original image in spatial domain. In our proposed method, soft shrinkage and Median Absolute Difference (MAD) are used. The scaled MAD noise estimator is calculated by (6).

$$MAD = \frac{median(|X|)}{0.6745} \tag{6}$$

where X is the high frequency sub-bands coefficients. From the estimated noise, the non linear threshold T is calculated by (7)

$$T = MAD * \sqrt{2logn} \tag{7}$$

where N is the size of the high frequency sub-band array. Then the soft thresholding is applied to remove the noise and the soft shrinkage rule is defined by (8). Finally, the noise free image is obtained by taking then inverse SWT

$$\rho_T(x) = \begin{cases} x - T, & if \quad x \geq T \\ x + T, & if \quad x \leq -T \\ 0, & if \quad |x| < T \end{cases} \tag{8}$$

The noise free sub-bands are obtained by using adaptive thresholding. Finally, the noise free image is obtained by taking the inverse SWT using the modified high frequencies sub-bands and the low frequency sub band of SWT.

IV. Experimental Results

The proposed algorithm tested for 256x256 images. It is tested for various levels of noise values and also compared with Standard median filter (SMF). Figure 2 shows the de-noising performance of the proposed algorithm. Table 2 show the PSNR values of the proposed method and based soft shrinkage and SWT method with different noise variance. Figure 3 shows the comparison of PSNR value for median filter and our proposed method.

(a) (b) (c) (d)

Figure 2 : (a) Input image, (b) Noise added Image, (c) Median filtered Image and (d) Our proposed method

Table 2 : PSNR value for the proposed method

Noise Level	Median Filter	Proposed Method
20	27.26	31.03
30	22.56	29.69
40	18.33	28.05
50	15.07	26.35
60	12.2	24.27
70	9.67	21.86

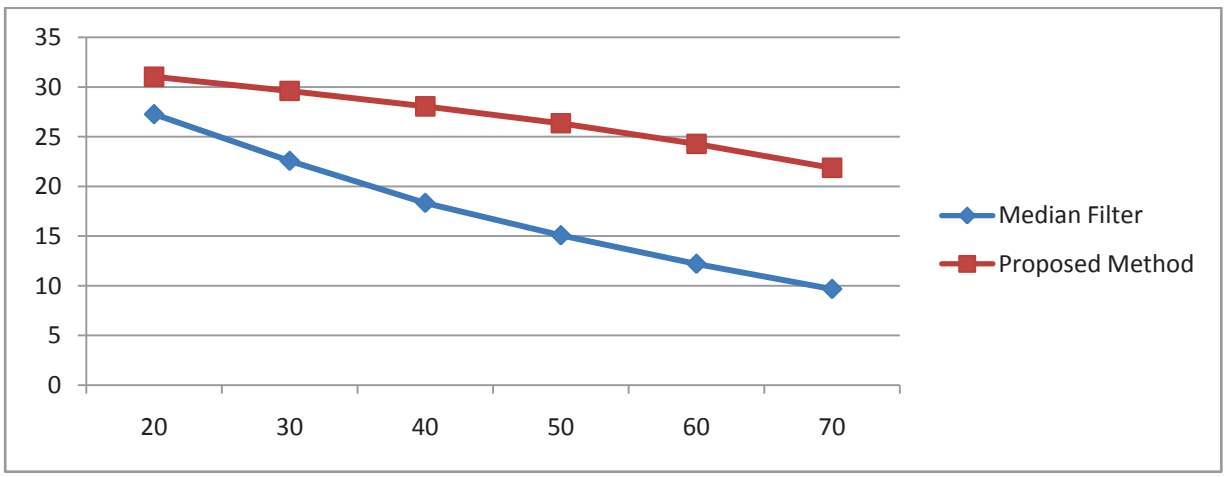


Figure 3 : Comparison graph of PSNR at different noise densities

V. Conclusion

In this work, we presented the image denoising based on Stationary Wavelet transform (SWT) and soft threshold method is presented. Experimental results show that the proposed method restore the original image much better than standard non linear median-based filters and some of the recently proposed algorithms. The proposed filter requires less computation time compared to other methods. The visual quality results clearly shows the proposed filter preserve fine details such as lines and corners satisfactorily. This filter can be further improved to apply for the images corrupted with high density impulse noise upto 90% and random valued impulse noise.

References Références Referencias

1. H.Hwang and R.A.Hadda, "Adaptive Median Filters: New Algorithms and Results", IEEE Transactions on Image Processing, 1995, pp 499-502.
2. Eduardo Abre, Michael Lightsone, "A New Efficient Approach for The Removal of Impulse Noise from Highly Corrupted Images", IEEE Transactions on Image Processing, 1996, pp 1012-1025.
3. Tao Chen, Kai-Kuang Ma, and Li-Hui Chen, "Tri-State Median Filter for Image Denoising" IEEE Transactions on Image Processing, 1999. Pp 1834-1838.
4. Zhou Wang and David Zhang, "Progressive Switching Median Filter for the Removal of Impulse Noise from Highly Corrupted Images" " IEEE Transactions on Circuits and Systems Analog And Digital Signal Processing, 1999. pp 78-80.
5. Tao Chen and Hong Ren Wu, "Space Variant Median Filters for the Restoration of Impulse Noise Corrupted Images" IEEE Transactions on Circuits and Systems Analog and Digital Signal Processing, 2001, pp 784-789.
6. How- Lung Eng and Kai-Kuang Ma, "Noise Adaptive Soft-Switching Median Filter", IEEE Transactions on Image Processing, 2001, pp 242-251.
7. Shuqun Zhang and Mohmamad A.Karim, "A New Fast and Efficient Decision- Based Algorithm for Removal of High- Density Impulse Noises", IEEE Signal Letter, 2002, pp 360-363.
8. Wenbin Luo, "An Efficient Detail- Preserving Approach for Removing Impulse Noise in Images", IEEE Signal Processing letter, 2006, pp 413-416.
9. K.S.Sreenivasan and D.Ebenezer, "A New Fast and Efficient Decision-Based Algorithm for Removal of High- Density Impulse Noises". IEEE Signal Letter, 2007, pp 189-192.
10. Dagao Duan, Quian Mo, "A Detail Preserving Filter for Impulse Noise removal", IEEE International Conference on Computer Application and System Modeling, 2010, pp 265-268

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 5 Version 1.0 March 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Handling of Congestion in Cluster Computing Environment Using Mobile Agent Approach

By P.K. Suri & Sumit Mittal
M.M. University, Mullana, Ambala, Haryana

Abstract - Computer networks have experienced an explosive growth over the past few years and with that growth have come severe congestion problems. Congestion must be prevented in order to maintain good network performance. In this paper, we proposed a cluster based framework to control congestion over network using mobile agent. The cluster implementation involves the designing of a server which manages the configuring, resetting of cluster. Our framework handles - the generation of application mobile code, its distribution to appropriate client, efficient handling of results, so generated and communicated by a number of client nodes and recording of execution time of application. The client node receives and executes the mobile code that defines the distributed job submitted by server and replies the results back. We have also the analyzed the performance of the developed system emphasizing the tradeoff between communication and computation overhead. The effectiveness of proposed framework is analyzed using JDK 1.5.

Keywords : *Cluster Computing, Mobile Agent, Congestion.*

GJCST Classification: *C.2.5*

Strictly as per the compliance and regulations of:

Handling of Congestion in Cluster Computing Environment Using Mobile Agent Approach

P.K. Suri [α] & Sumit Mittal [σ]

Abstract - Computer networks have experienced an explosive growth over the past few years and with that growth have come severe congestion problems. Congestion must be prevented in order to maintain good network performance. In this paper, we proposed a cluster based framework to control congestion over network using mobile agent. The cluster implementation involves the designing of a server which manages the configuring, resetting of cluster. Our framework handles - the generation of application mobile code, its distribution to appropriate client, efficient handling of results, so generated and communicated by a number of client nodes and recording of execution time of application. The client node receives and executes the mobile code that defines the distributed job submitted by server and replies the results back. We have also the analyzed the performance of the developed system emphasizing the tradeoff between communication and computation overhead. The effectiveness of proposed framework is analyzed using JDK 1.5.

Keywords : Cluster Computing, Mobile Agent, Congestion.

I. Introduction

Mobile agent technology plays a vital role in the management of cluster computing. Mobile agents are autonomous programs that travel from computer to computer under their own control. Their abilities to cope with system heterogeneity and to deploy user-customized procedures at remote sites are well suited for cluster computing environments[12]. By interacting with a remote host after migrating to it, an agent can perform complex operations on remote data without transferring them, because the agent can carry the application logic to where it is needed.

Cluster computing harnesses the combined computing power of multiple microprocessors in a parallel configuration. Cluster computers are a set of commodity PC's dedicated to a network designed to capture their cumulative processing power for running parallel-processing applications. Clustere computers are specifically designed to take large programs and sets of data and subdivide them into component parts, thereby allowing the individual nodes of the cluster to process their own individual chunks of the program.

Author [α] : Dean (R&D) & Chairman (CSE/IT/MCA), HCTM Technical Campus, Kaithal, Haryana, 136 027, India
E-mail : pksuritf25@yahoo.com
Author [σ] : M.M. Institute of Computer Technology & Business Management, M.M. University, Mullana, Ambala, Haryana, 133 207, India E-mail : sumit_amb@yahoo.com

Literature Survey : A number of research efforts in the area of improving the performance of distributed applications in a cluster computing environment have emerged[2][7][8]. Martin, Vahdat, Culler and Anderson[9] revealed that the communication latency, overhead, and bandwidth can be independently varied to observe the effects on a wide range of applications. They showed that applications demonstrate strong sensitivity to overhead, slowing down, when overhead is increased. Lange, D.B[10] discussed the impact of cluster size and underlying network on the performance of distributed applications.

Our research effort (i) evaluates the performance of application as a function of problem size, network parameters and cluster size (ii) provides features for monitoring of cluster. The developed MCLUSTER model concentrates on reducing the communication overhead by using the mobility of code and bridging all system characteristics.

II. Architecture of Cluster Computing

There are many cluster configurations, but a simple architecture is shown in the Figure 1.

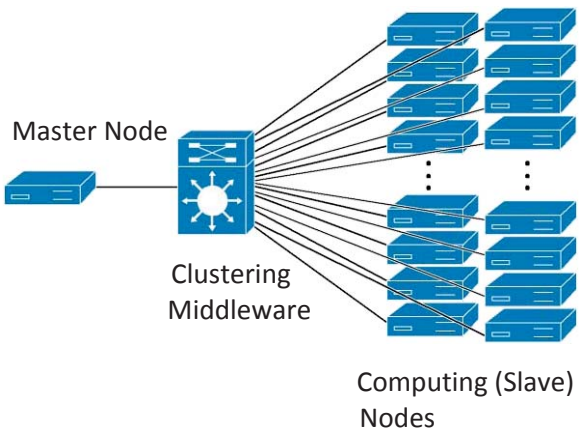


Figure 1. Cluster Computing Architecture

As shown in Figure 1, node participation in the cluster falls into master or head node and computing or slave nodes. The master node is the unique server in cluster systems. It is responsible for running the file system and also serves as the key system for clustering middleware to route processes, duties and monitor the

status of each slave node. A slave node within a cluster provides the cluster a computing and data storage capability. These nodes are derived from fully operational, standalone computers that are typically marketed as desktop or server systems that are off-the-shelf commodity systems.

III. Design of Cluster Computing

The framework consists personal computers, high speed communication network, sequential and parallel applications as shown in Figure 2.

PC's are connected to the network using Ethernet Network Interface Card. Cluster middleware is implemented in JAVA so that middleware can provide the single system image of the cluster to any computer with different OS platforms, once the JAVA virtual machine is installed. JVM makes it easier to implement, migrate and execute the mobile code at remote computer in the cluster. The user is guided through the creation and management of cluster via GUI. It frees the user from identifying the network topology of the framework of cluster. The framework has been designed in such a way that incremental changes to it can easily enhance the generality and usability of cluster.

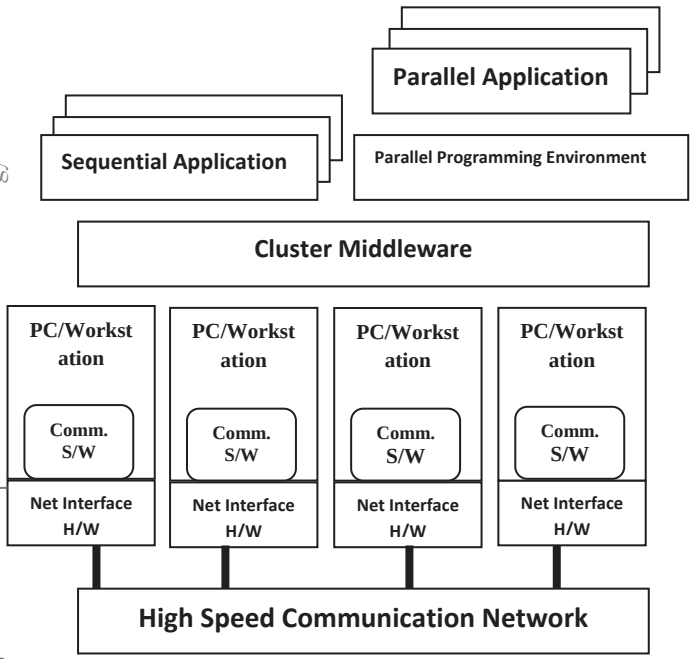


Figure 2. Cluster Computing Framework

IV. Cluster Middleware Architecture

Cluster middleware allows the user to develop & process the parallel applications on clusters and achieve good performance. In cluster configuration, there are two types of nodes: a server node named MCLUSTER, client nodes as shown in the Figure 3.

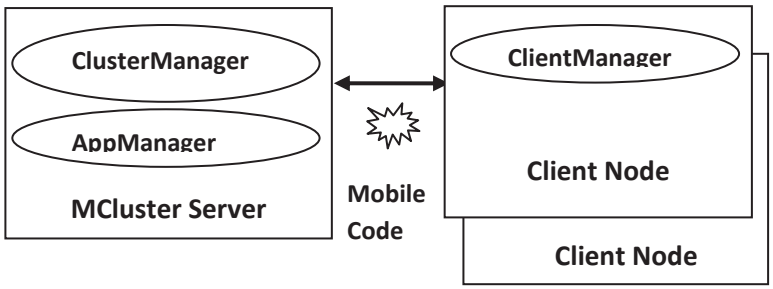


Fig. 3 Cluster middleware

MCLUSTER and client nodes are interacting with each other through message passing and any parallel application is executed using mobile code which contains input data/application code.

V. ClusterManager

ClusterManager initiates the cluster and controls all nodes of the cluster. ClusterManager periodically broadcast invitation packets to all nodes in the network as shown in the Figure 4. All those nodes on which ClientManager is activated and are not being part of the cluster, will display a frame on the user screen after receiving the invitation packet. If client node accepts the invitation, then, IP address of that node will be automatically communicated to MCLUSTER as show in Figure 4. MCLUSTER maintains a database of all these addresses.

VI. AppManager

AppManager coordinates all functions related to application description, distribution and execution. AppManager provides a GUI from which user can easily select a particular application, provides the input data and specify the parameters such as number of nodes to be used for the execution of application. Once the application is selected, then the AppManager will divides the data range into the number of nodes and distribute the mobile code to the selected client nodes as shown in Figure 4.

VII. ClientManager

ClientManager listens to the requests from MCLUSTER related to cluster membership, application execution and service those requests. In lieu of invitation packet, ClientManager running on a node will communicate its IP address to the server. On receiving the mobile code from server it load and execute the mobile code and reply the results back.

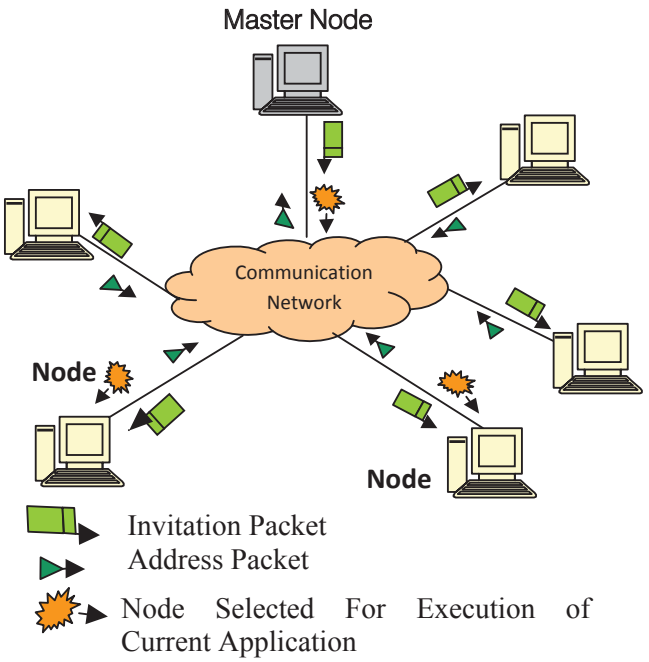


Figure 4 Communication between Master node and client node

VIII. Implementation of Our Framework

In our research effort, series application is selected for validating the framework. It basically involves generating a sequence of numbers between an initial value & a final value. MCLUSTER begins a timer before the execution of application and stops it after the collection of entire series from client nodes as shown in the figure 5. The effect of number of client nodes and data range on the execution time of application is represented by Figure 6 & 7.

The performance of cluster, during the execution of distributed applications not having significant amount of data (such as between 1-20000) deteriorate as indicated by Overhead point in Figure 6. At these data ranges the communication overhead between MCLUSTER Server and Client nodes overwhelm the advantage of distributed processing power obtained from client nodes. When data range is extensive then the time consumed by an application so executed in a distributed manner is several times smaller than the execution time of the same application processed on a single node. The MCLUSTER model is designed in such a way that even for a large problem size (such as between 1-5000000) and average cluster size (up to N=11) performance of distributed application will not deteriorate as shown in Figure 7.

Communication Overhead Analysis : When the scalability of cluster increases, the communication links near the server are congested due to large transmission of data, thereby degrading the performance of cluster. Communication overhead includes the overhead due to exchange of data between nodes and message delay caused by network congestion. Figure 8 is represented by the result so generated after execution of the application Series for a data range 1-100000 on a cluster of 11 client nodes.

On the basis of assumption that external network load is low and constant, the communication overhead for MCLUSTER model can be parameterized as a simple linear function of number of bytes transmitted and number of client nodes selected as represented in Figure 8.

$$\text{Coverhead} = T + C * N$$

Where = Startup time involving Partitioning time, Time spent during selection of client nodes.
C = Cost per byte transmitted.
N = Number of nodes selected for a particular execution.

IX. Conclusion

Cluster computing undoubtedly is gaining importance as a substitute for very expensive parallel computers. As depicted in our research work, Cluster computing by appropriately combining the computational processing power of autonomous computers can significantly improve the performance of distributed applications. Cluster Computing on the other hand can decline the performance of a problem if a proper check is not applied at the problem size as well as cluster size. Since the improper values of these two parameters will lead to the network congestion thereby resulting in the overwhelming of computation load by communication load. However this Network Congestion problem can be easily tackled by using the concept of Mobile Agent.

References Références Referencias

1. Baker, M. and R. Buya, 1999. Cluster computing: The commodity supercomputer. Software-Prentice, 29: 551-576.
2. Cluster Computing Overview, http://www.cs.wayne.edu/~haj/load/survey/overview.html.
3. Becker, D. and P. Merkey. The Beowulf Project. http://www.beowulf.org.
4. Message Passing Interface, MPI Forum-http://www.mpi-forum.org.
5. Flynn, M.J. Very High Speed Computing Systems. Proceedings. IEEE, 54: 1901.
6. Geist,A. and J. Schwidder, 1999. Managing multiple multi-user PC clusters. J. Parallel and Distributed Computing.
7. Cheung, L. and A.P. Reeves, 1992. High performance computing on a cluster of workstations. Proceedings of First Symposium on High - Performance Distributed Computing.

8. Jon, B.W. and A.S. Grimshaw, 1994. Network partitioning of data parallel computations. Proc. Third International Symposium. High Performance Distributed Computing (HPDC '94), April 2-5, 1994, San Francisco, CA, USA. IEEE Computer Society.
9. Lemeire, J. and E. Dirkx, 2002. Causes of blocking overhead in message-passing programs.10th Euro PVM/MPI 2002 Conference, Venice, Italy.
10. Martin, R., A. Vahdat, D. Culler and T. Anderson, 1997. Effects of communication latency,overhead and bandwidth in a cluster architecture. Proceedings 24th Annual International. Symposium Computer Architecture (ISCA), New York, USA, pp: 85-97.11. Lange, D.B. Mobile objects and Mobile Agents: The Future of Distributed Computing? General Magic, Inc. California.
11. Walker, B. and D. Steel, 1999. Implementing a full single system image unixware cluster: Middleware vs. underware. Proceedings: International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'99), Las Vegas, USA.
12. Ichiro Satoh. Configurable Network Processing for Mobile Agents on the Internet. Cluster computing, The Journal of Networks, Software Tools and Applications), vol. 7, no.1, pp.73-83, Kluwer, 2004.

Data Range	No. of Nodes	Time (Milliseconds)
1-10000	1	1718
	2	1469
	3	1282
1-20000	1	3000
	2	1734
	3	1890
1-50000	1	7016
	2	3734
	3	3125
1-100000	1	13687
	2	7015
	3	5375

Figure 5. Statistical Data of Application : Series

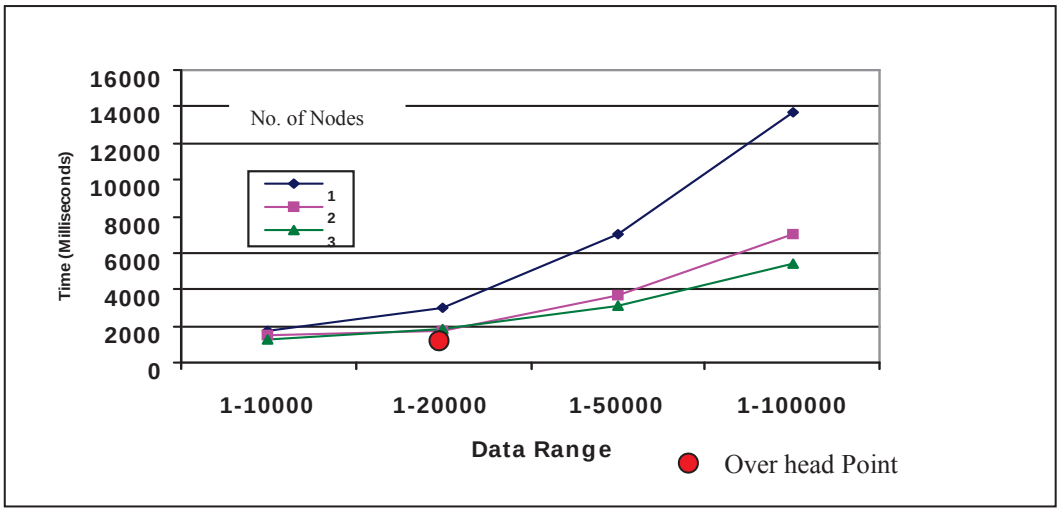


Figure 6. Effect of data range and number of nodes on execution time

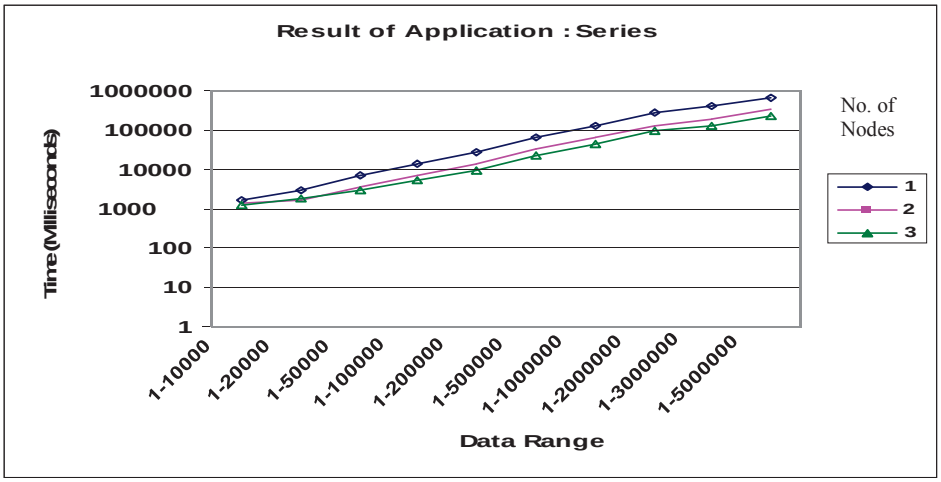


Figure 7. Effect of data range on execution time of application series

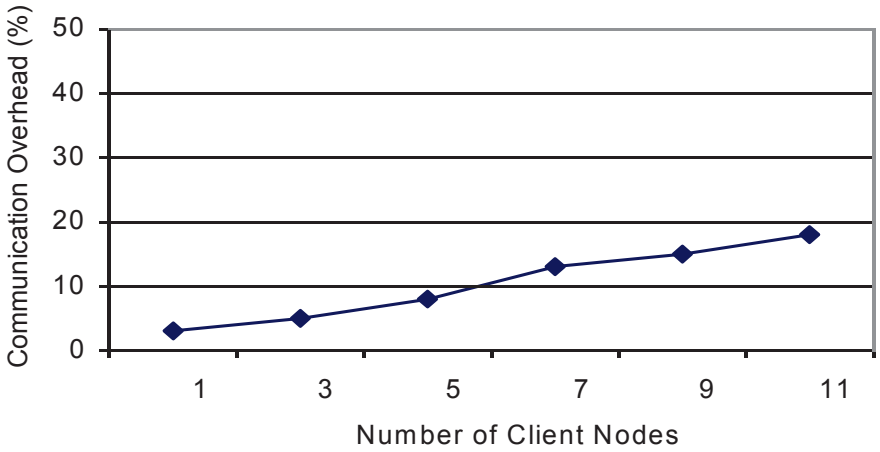


Figure 8. Communication overhead vs. cluster size

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 5 Version 1.0 March 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Breaking of Simplified Data Encryption Standard Using Genetic Algorithm

By Lavkush Sharma, Bhupendra Kumar Pathak & Ramgopal Sharma
Fet Rbs College ,Agra

Abstract - Cryptanalysis of ciphertext by using evolutionary algorithm has gained so much interest in recent years. In this paper we have used a Genetic algorithm with improved crossover operator (Ring Crossover) for cryptanalysis of S-DES. There so many attacks in cryptography. The cipher text attack only is considered here and several keys are generated in the different run of the genetic algorithm on the basis of their cost function value which depends upon frequency of the letters. The results on the S-DES indicate that, this is a promising method and can be adopted to handle other complex block ciphers like DES, AES.

Keywords : *Cryptanalysis, Ciphertext attack, Simplified Data Encryption Standard, genetic algorithm, Key search space*

GJCST Classification: *E.3*

Strictly as per the compliance and regulations of:

Breaking of Simplified Data Encryption Standard Using Genetic Algorithm

Lavkush Sharma [α], Bhupendra Kumar Pathak [σ] & Ramgopal Sharma [ρ]

Abstract - Cryptanalysis of ciphertext by using evolutionary algorithm has gained so much interest in recent years. In this paper we have used a Genetic algorithm with improved crossover operator (Ring Crossover) for cryptanalysis of S-DES. There so many attacks in cryptography. The cipher text attack only is considered here and several keys are generated in the different run of the genetic algorithm on the basis of their cost function value which depends upon frequency of the letters. The results on the S-DES indicate that, this is a promising method and can be adopted to handle other complex block ciphers like DES, AES.

Keywords : *Cryptanalysis, Ciphertext attack, Simplified Data Encryption Standard, genetic algorithm, Key search space*

I. Introduction

A cipher is a secret way of writing in which plaintext is converted into a scrambled (encrypted) version of the original message (ciphertext) by using a key. Those who know the key can easily decrypt the ciphertext back into the plaintext. Cryptanalysis is the study of breaking ciphers that is finding the key or converting the ciphertext into the plaintext without knowing the key. Many cryptographic systems have a finite key space and, hence, are vulnerable to an exhaustive key search attack. Yet, these systems remain secure from such an attack because the size of the key space is such that the time and resources for a search are not available. Optimization techniques have got a significant importance in determining efficient solutions of different complex problems. One such problem is to break S-DES. This paper considers cryptanalysis of S-DES. In the brute force attack, the attacker tries each and every possible key on the part of cipher text until desired plaintext is obtained. A brute force approach may take so much time to guess the real key which is used to generate a cipher text, so the difficulty of breaking the cipher is directly proportional to the number of keys. On the other hand optimization technique can be used for the same purpose. Genetic algorithm is an evolutionary algorithm that works well and takes less time to break cipher as compared to Brute force attack.

Author [α] : LavkushSharma,FET RBS Agra,
E-mail : lavkush07@yahoo.com
Author [σ] : B. K. Pathak, juit, solan,
E-mail : bhupendra.pathak@juit.ac.in
Author [ρ] : R.G Sharma, FET RBS Agra,
E-mail : cs.ramgopal@gmail.com

The remaining paper is organized as follows: Section II discusses the earlier works done in this field. Section III presents overview of S-DES and Section IV gives the overview of Genetic Algorithm. Experimental results are discussed in Section V.Conclusion are presented in section VI At last References are given.

II. Related Work

In the last few years, so many papers have been published in the field of cryptanalysis. R.Spillman etc. showed that Knapsack cipher[4] and substitution ciphers[5] could be attacked using genetic algorithm. In the recent years Garg[1,2] presented the use of memetic algorithm and genetic algorithm to break a simplified data encryption standard algorithm. Nalini[3] used efficient heuristics to attack S-DES. In 2006 Nalini used GA, Tabu search and Simulated Annealing techniques to break S-DES. Matusi[7] showed the first experimental cryptanalysis of DES using an linear cryptanalysis technique. Clark[6] also presented important analysis on how different optimization techniques can be used in the field of cryptanalysis. Vimalathithan[9] also used GA to attack Simplified-DES. In this paper, a Genetic Algorithm with improved parameters is used to break S-DES.A population of keys is generated and their fitness is calculated by using efficient fitness function. At the end, we will find the key in less time.

III. S-Des

In this section we will provide the overview of S-DES Algorithm. Simplified DES, developed by Professor Edward Schaefer of Santa Clara University is an educational rather than a secure encryption algorithm. The S-DES [8, 10] encryption algorithm takes an 8-bit block of plaintext and a 10-bit key as input and produces an 8-bit block of ciphertext as output. The S-DES decryption algorithm takes an 8-bit block of ciphertext and the same 10-bit key used to produce that ciphertext as input and produces the original 8-bit block of plaintext. The encryption algorithm involves five functions: an initial permutation (IP); a complex function labeled f_K, which involves both permutation and substitution operations and depends on a key input; a simple permutation function that switches (SW) the two halves of the data; the function f_K again; and finally a

permutation function that is the inverse of the initial permutation (IP^{-1}).

The function f_K takes as input not only the data passing through the encryption algorithm, but also an 8-bit key. S-DES uses a 10-bit key from which two 8-bit subkeys are generated. In this, the key is first subjected to a permutation (P10). Then a shift operation is performed. The output of the shift operation then passes through a permutation function that produces an 8-bit output (P8) for the first subkey (K_1). The output of the shift operation also feeds into another shift and another instance of P8 to produce the second subkey (K_2).

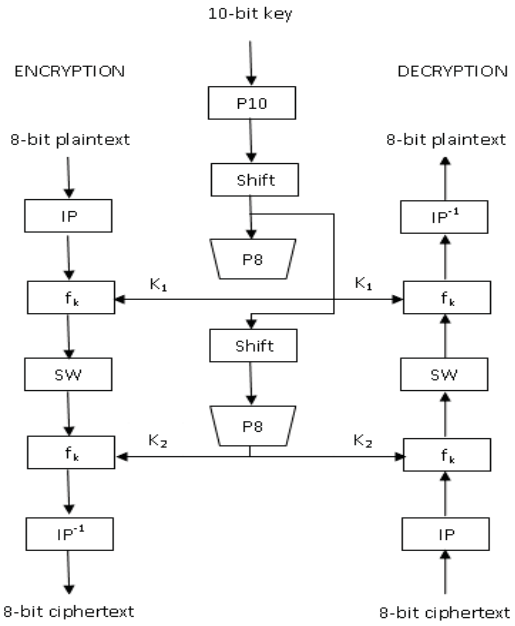


Figure 1: Simplified Data Encryption Algorithm

a) Initial and Final Permutations

The input to the algorithm is an 8-bit block of plaintext, which we first permute using the IP function IP= [2 6 3 1 4 8 5 7].This retains all 8-bits of the plaintext but mixes them up. At the end of the algorithm, the inverse permutation is applied; the inverse permutation is done by applying, IP^{-1} = [4 1 3 5 7 2 8 6] where we have $IP^{-1}(IP(X))$ =X.

b) The Function f_K

The function f_k, which is the complex component of S-DES, consists of a combination of permutation and substitution functions. The functions are given as follows.

Let L, R be the left 4-bits and right 4-bits of the input, then,

$$f_K (L, R) = (L \text{ XOR } f(R, key), R)$$

Where XOR is the exclusive-OR operation and key is a sub -key. Computation of f(R, key) is done as follows.

1. Apply expansion/permutation E/P= [4 1 2 3 2 3 4 1] to input 4-bits.
2. Add the 8-bit key (XOR).
3. Pass the left 4-bits through S-Box S0 and the right 4-bits through S-Box S1.
4. Apply permutation P4 = [2 4 3 1].

$$S0 = \begin{array}{c|cccc} & 0 & 1 & 2 & 3 \\ \hline 0 & 1 & 0 & 3 & 2 \\ 1 & 3 & 2 & 1 & 0 \\ 2 & 0 & 2 & 1 & 3 \\ 3 & 3 & 1 & 3 & 2 \end{array}$$

$$S1 = \begin{array}{c|cccc} & 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 1 & 2 & 3 \\ 1 & 2 & 0 & 1 & 3 \\ 2 & 3 & 0 & 1 & 0 \\ 3 & 2 & 1 & 0 & 3 \end{array}$$

Figure 2 : Working of S-box

The S-boxes operate as follows:

The first and fourth input bits are treated as 2-bit numbers that specify a row of the S-box and the second and third input bits specify a column of the Sbox.

The entry in that row and column in base 2 is the 2-bit output.

c) The Switch Function

The function f_K only alters the leftmost 4 bits of the input. The switch function (SW) interchanges the left and right 4 bits so that the second instance of f_K operates on a different 4 bits. In this second instance, the E/P, S0, S1, and P4 functions are the same. The key input is K2.

IV. Genetic Algorithm

The genetic algorithm [13, 20] is a search algorithm based on the natural selection and on "survival of the fittest", the main idea is that in order for a population of individuals to adapt to some environment, it should behave like a natural system. This means that survival and reproduction of an individual is promoted by the elimination of useless traits and by rewarding useful behavior. The genetic algorithm belongs to the family of evolutionary algorithms. An evolutionary algorithm maintains a population of solutions for the problem at hand. The population is then evolved by the iterative application of a set of stochastic operators. The simplest form of genetic algorithm involves three types of operators: selection, crossover and mutation.

A selection operator is applied first.

Selection : This selection operator selects chromosomes in the population for reproduction. The better the

chromosome, the more times it is likely to be selected to reproduce.

Crossover : Crossover selects genes from is parent chromosomes and creates a new offspring. The simplest way to do this is to choose randomly some crossover point and everything before this point is copied from the first parent and then, everything after a crossover point copied from the second parent.

Mutation : After a crossover, mutation is performed. This is to prevent falling all solutions in population into a local optimum of solved problem. Mutation changes randomly the new offspring. In binary GA we can switch a few randomly chosen bits from 1 to 0 or from 0 to 1.

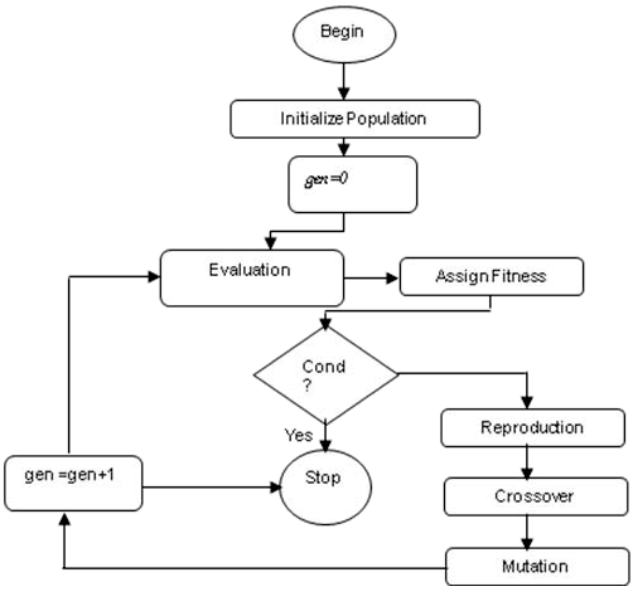


Figure 3 : Flow chart of Genetic Algorithm [19]

In this paper, we are using Ring crossover operator [11].In ring crossover two parents such as parent1 and parent2 are considered for the crossover process, and then combined in the form of ring, as shown in fig. 4 (b). Later, a random cutting point is decided in any point of ring. The children are created with a random number generated in any point of ring according to the length of the combined two parental chromosomes. With reference to the cutting point, while one of the children is created in the clockwise direction, the other one is created in direction of the anti-clockwise, as shown in fig. 4(c). Then swapping and reversing process is performed in the Ring Crossover operator, as shown in fig. 4(d).

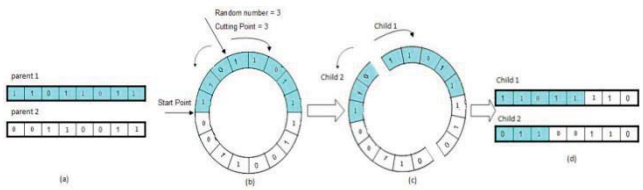


Figure 4 : Ring Crossover Procedure [11]

The primary goals of this work are to produce a performance comparison between traditional Brute force search algorithm and genetic algorithm with improved parameters based method, and to determine the use of typical GA-based methods in the field of cryptanalysis.

The procedure to carry out the cryptanalysis using GA in order to break the key is as follows

1. Input: ciphertext, and the language statistics.
2. Randomly generate an initial pool of solutions (keys).
3. Calculate the fitness value of each of the solutions in the pool using equation (1).
4. Create a new population by repeating following steps until the new population is complete

a. Select parent (keys) from a current population according to their fitness value (the better fitness, the bigger chance to be selected). Here Tournament selection is used.
b. With a crossover probability cross over the parents to form new offspring (childrens). In our genetic algorithm we are using Ring Crossover Operator
c. For each of the children, perform a mutation operation with some mutation probability to generate new children.
d. Place new children in the new population

5. Use new generated population for a further run of the algorithm
6. If the end condition is satisfied, stop, and return the best solution in current population

a) Cost Function

Equation (1) is a general fitness function used to determine the suitability of a assumed key (k). Here, A denotes the language alphabet (i.e., for English, [A... Z, _], where _ represents the space symbol), K and D denote known language statistics and decrypted message statistics, respectively, and the u, b, and t denote the unigram, digram and trigram statistics respectively; α, β and γ are the weights assigning different weights to each of the three statistics where $\alpha + \beta + \gamma = 1$. In view of the computational complexity of trigram, only unigram and digram statistics are used.

$$C^K = \alpha \ \Sigma(i \varepsilon \tilde{A}) \ |K(i)^u - D(i)^u| + \beta \ \Sigma(i, j \varepsilon \tilde{A}) \ | \ K(i, j)^b - D(i, j)^b| + \gamma\Sigma(i,j,k\varepsilon\tilde{A}) \, |K(i,j,k)^t - D(i,j,k)^t| \qquad (!)$$

V. Results and Discussion

Our objective in this paper is to compare the results obtained from Brute Force search algorithm with the Genetic Algorithms with improved parameters. The experiments were conducted on Core 2 Duo system. There are a variety of cost functions used by other researchers in the past. The most common cost function uses gram statistics. Some use a large amount of grams

while others only use a few. Equation 1 is a general formula used to determine the suitability of a proposed key. A number of experiments have been carried out by giving different inputs and applying genetic algorithm and Brute force attacks for breaking Simplified Data Encryption Standard. The results are shown in table 1.The table below shows that the key bits matched using GA and Brute Force search algorithm for the given cipher text .the choice of the Genetic Operators play a vital role in GA and are described below:

GA Parameters

The following are the GA parameters used during the experimentation:

Population Size: 100
Selection : Tournament Selection operator
Crossover Ring Crossover
Crossover: .85
Mutation: .02
No. of Generation: 50

Table 1: Comparison of Genetic Algorithm and Brute Force Search Algorithm.

S. No	Amount of Cipher Text	No. of bits matched using GA	No. of bits matched using Brute Force search	Time Taken by GA (M)	Time Taken by Brute Force search (M)
1.	200	5	5	4.7	24.3
2.	400	4	3	2.1	24.7
3.	600	7	6	1.9	23.6
4.	800	8	7	3.1	24.1
5.	1000	9	7	2.6	25.1
6.	1200	9	8	2.1	25.5

From the above table, it is found that both GA works better than Brute force algorithm in terms of time taken as well as obtaining number of key bits.

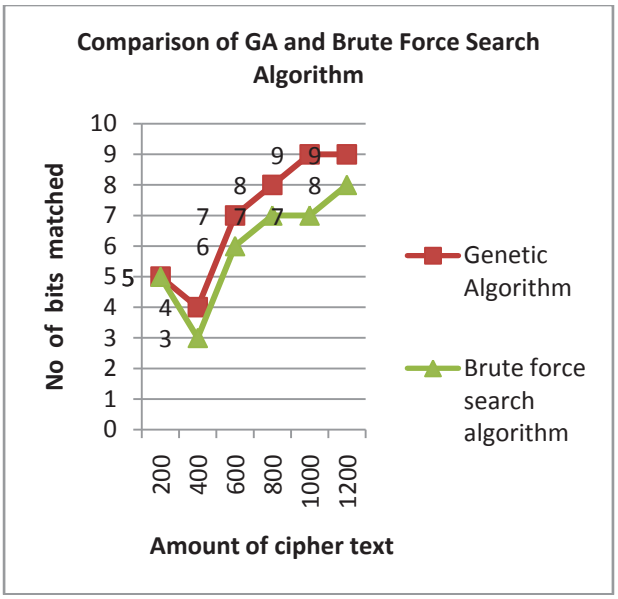


Figure 5 : comparison of Genetic algorithm and Brute Force Search algorithm

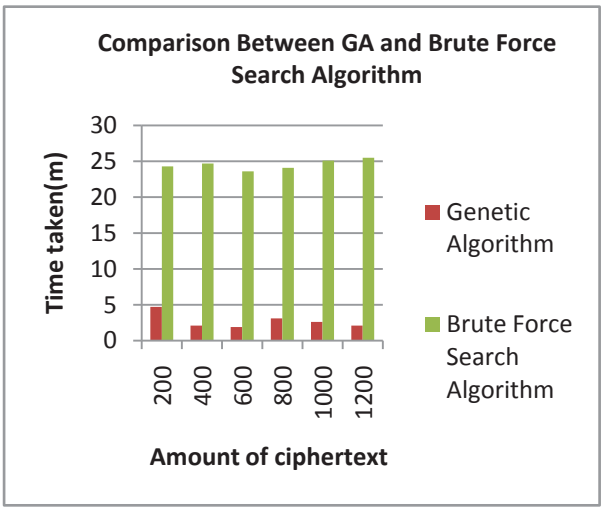


Figure 6 : The running time comparison of Genetic Algorithm and Brute Force Search Algorithm

VI. Conclusion

In this paper, we have used a Genetic algorithm with Ring crossover and other operators for the cryptanalysis of Simplified Data Encryption Standard. We found that Genetic Algorithm is far better than Brute Force search algorithm for cryptanalysis of S-DES. Although S-DES is a simple encryption algorithm, GA with Ring Crossover method can be adopted to handle other complex block ciphers like DES and AES.

References Références Referencias

1. G Poonam, Memetic Algorithm Attack on Simplified Data Encryption Standard algorithm, proceeding of International Conference on Data Management, February 2008, pg 1097-1108

2. Garg Poonam, Genetic algorithm Attack on Simplified Data Encryption Standard Algorithm, International journal Research in Computing Science,ISSN1870-4069, 2006.
3. Nalini, Cryptanalysis of S-DES via Optimization heuristics, International Journal of Computer Sciences and network security, vol 6, No 1B, Jan 2006.
4. Spillman, R.: Cryptanalysis of Knapsack Ciphers Using Genetic Algorithms.Cryptologia XVII(4), 367–377 (1993)
5. Spillman, R., Janssen, M., Nelson, B., Kepner, M.: Use of A Genetic Algorithm in the Cryptanalysis of simple substitution Ciphers. Cryptologia XVII(1), 187–201 (1993)
6. Clark A and Dawson Ed, "Optimisation Heuristics for the Automated Cryptanalysis of Classical Ciphers", Journal of Combinatorial Mathematics and Combinatorial Computing, Vol.28,pp. 63-86, 1998.
7. M. Matsui, Linear cryptanalysis method for DES cipher, Lect. Notes Comput. Sci. 765 (1994) 386–397.
8. William Stallings, Cryptography and Network Security Principles and Practices, Third Edition,Pearson Education Inc.,2003.
9. Vimalathithan.R, M.L.Valarmathi, "Cryptanalysis of SDES Using Genetic Algorithm", International Journal of Recent Trends in Engineering, Vol2, No.4, November 2009, pp.76-79.
10. Schaefer E, "A Simplified Data Encryption Standard Algorithm", Cryptologia, Vol .20, No.1, pp. 77-84, 1996.;
11. Yılmaz Kaya, Murat Uyar, Ramazan Tekdn," A Novel Crossover Operator for Genetic Algorithms: Ring Crossover".
12. Davis,L. "Handbook of Genetic Algorithm",Van Nostrand Reinhold, New York,1991
13. D. E. Goldberg,"Genetic algorithms in search. Optimization and Machine Learning.Reading. M.A. addison -Wesley.1989.
14. A,Michalewiez and N. Attia." Evolutionary optimization of constrained problems." InProc.3rd annu. Conf. on Evolutionary Programming. 1994.pp 98-108
15. Z. Michalewicz. "Genetic algorithms+ Data structures = Evolution programs 3rd ed. New York. springer,1996.
16. N.Koblitz, "A Course on number theory and cryptography", Springer-Verlag New York,Inc., 1994.
17. Alfred J. Menezes. Menezes, Alfred J. Handbook of Applied Cryptography, CRC, 1997.
18. R. Toemeh, S. Arumugam, Breaking Transposition Cipher with Genetic Algorithhm Electronics and Electrical Engineering,ISSN 1392 – 1215 2007. No. 7(79)
19. Kalyanmoy Deb, Multi-objective Optimization using Evolutionary Algorithms, John Wiley and Sons, 2001.
20. C.W. Wu and N. F. Rulkov, —Studying chaos via 1-Dmaps—atutorial,‖ IEEE Trans. on Circuits and Systems I: Fundamental Theory and Applications, vol. 40, no. 10, pp. 707–721, 1993.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Volume 12 Issue 5 Version 1.0 March 2012
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Performance and Effectiveness of Secure Routing Protocols in Manet

By Er.Gurjeet Singh
Desh Bhagat Institute of Engg & Management Moga

Abstract - Mobile adhoc network (MANET) is a temporary network setup for a specific purpose without help of any preexisting infrastructure. The nodes in MANET are empowered to exchange packet using a radio channel. The nodes not in direct reach of each other uses their intermediate nodes to forward packets. (MANET) environment of MANET makes it vulnerable to various network attacks. A common type of attacks targets at the underlying routing protocols. Malicious nodes have opportunities to modify or discard routing information or advertise fake routes to attract user data to go through themselves. Some new routing protocols have been proposed to address the issue of securing routing information. However, a secure routing protocol cannot singlehandedly guarantee the secure operation of the network in every situation. The objectives of the paper is to study the performance and effectiveness of some secure routing protocols in these simulated malicious scenarios, including ARIADNE and the Secure Ad hoc On-demand Distance Vector routing protocol (SAODV).

GJCST Classification: C.2.2

Strictly as per the compliance and regulations of:

Performance and Effectiveness of Secure Routing Protocols in Manet

Er.Gurjeet Singh

Abstract - Mobile adhoc network (MANET) is a temporary network setup for a specific purpose without help of any pre-existing infrastructure. The nodes in MANET are empowered to exchange packet using a radio channel. The nodes not in direct reach of each other uses their intermediate nodes to forward packets. (MANET) environment of MANET makes it vulnerable to various network attacks. A common type of attacks targets at the underlying routing protocols. Malicious nodes have opportunities to modify or discard routing information or advertise fake routes to attract user data to go through themselves. Some new routing protocols have been proposed to address the issue of securing routing information. However, a secure routing protocol cannot single-handedly guarantee the secure operation of the network in every situation. The objectives of the paper is to study the performance and effectiveness of some secure routing protocols in these simulated malicious scenarios, including ARIADNE and the Secure Ad hoc On-demand Distance Vector routing protocol (SAODV).

I. Introduction

Mobile Ad-hoc network is a set of wireless devices called wireless nodes, which dynamically connect and transfer information. Wireless nodes can be personal computers (desktops/laptops) with wireless LAN cards, Personal Digital Assistants (PDA), or other types of wireless or mobile communication devices. Figure 1 illustrates what MANET is. In general, a wireless node can be any computing equipment that employs the air as the transmission medium. As shown, the wireless node may be physically attached to a person, a vehicle, or an airplane, to enable wireless communication among them..

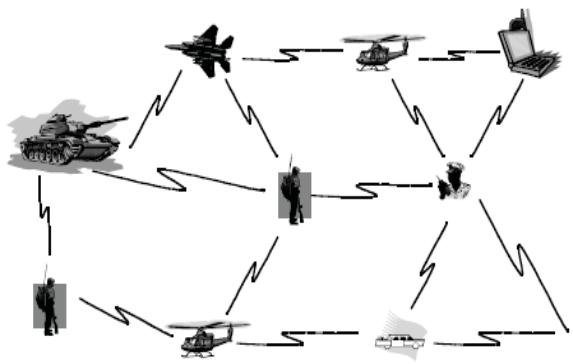
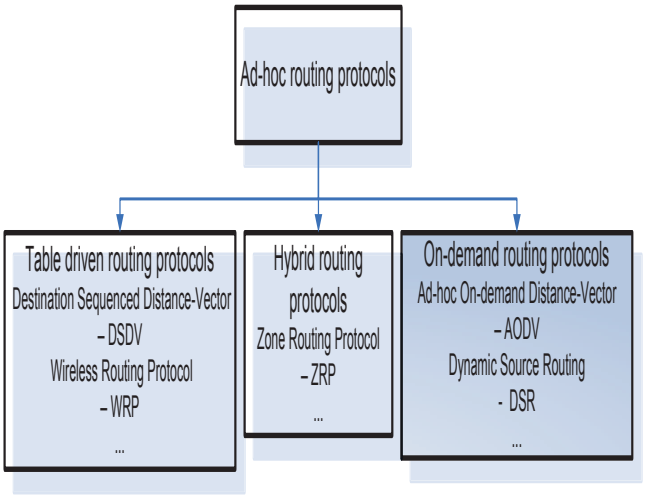


Figure1: Mobile Ad-hoc Network

Author : AP, Deptt of CSE Desh Bhagat Institute of Engg & Management Moga

II. Adhoc Wireless Routing Protocols

Routing protocols in ad hoc mobile wireless network can generally be divided into three groups



- **Table driven:** Every node in the network maintains complete routing information about the network by periodically updating the routing table. Thus, when a node needs to send data packets, there is no delay for discovering the route throughout the network. This kind of routing protocols roughly works the same way as that of routing protocols for wired networks.
- **Source initiated (or demand driven):** In this type of routing, a node simply maintains routes to active destination that it needs to send data. The routes to active destinations will expire after some time of inactivity, during which the network is not being used.
- **Hybrid:** This type of routing protocols combines features of the above two categories. Nodes belonging to a particular geographical region or within a certain distance from a concerned node are said to be in the routing zone and use table driven routing protocol. Communication between nodes in different zones will rely on the on-demand or source-initiated protocols.

III. Dynamic Source Routing Protocol(Dsr)

The Dynamic Source Routing Protocol is one of the on-demand routing protocols, and is based on the concept of *source routing*. In source routing, a sender node has in the packet header the complete list of the path that the packet must travel to the destination node. That is, every node in the path just forwards the packet to its next hop specified in the header without having to check its routing table as in table-driven routing protocols. Besides, the nodes don't have to periodically broadcast their routing tables to the neighboring nodes. This saves a lot of network bandwidth. The two phases of the DSR operation are described below:

- Route Discovery phase

In this phase, the source node searches a route by broadcasting route request (RREQ) packets to its neighbors. Each of the neighbor nodes that has received the RREQ broadcast then checks the packet to determine which of the following conditions apply: (a) Was this RREQ received before ? (b) Is the TTL (Time To Live) counter greater than zero? (c) Is it itself the destination of the RREQ? (d) Should it broadcast the RREQ to its neighbors? The *request ids* are used to determine if a particular route request has been previously received by the node. Each node maintains a table of RREQs recently received. Each entry in the table is a *<initiator, request id>* pair. If two RREQs with the same *<initiator, request id>* are received by a node, it broadcasts only the one received first and discards the other. This mechanism also prevents formation of routing loops within the network. When the RREQ packet reaches the destination node, the destination node sends a reply packet (RREP) on the reverse path back to the sender. This RREP contains the recorded route to that destination.

Figure 2 shows an example of the route discovery phase. When node A wants to communicate with node G, it initiates a route discovery mechanism and broadcasts a request packet (RREQ) to its neighboring nodes B, C and D as shown in the figure. However, node C also receives the same broadcast packets from nodes B and D. It then drops both of them and broadcasts the previously received RREQ packet to its neighbors. The other nodes follow the same procedure. When the packet reaches node G, it inserts its own address and reverses the route in the record and unicasts it back on the reversed path to the destination which is the originator of the RREQ.

The destination node unicasts the best route (the one received first) and caches the other routes for future use. A *route cache* is maintained at every node so that, whenever a node receives a route request and finds a route for the destination node in its own cache, it sends a RREP packet itself instead of broadcasting it further.

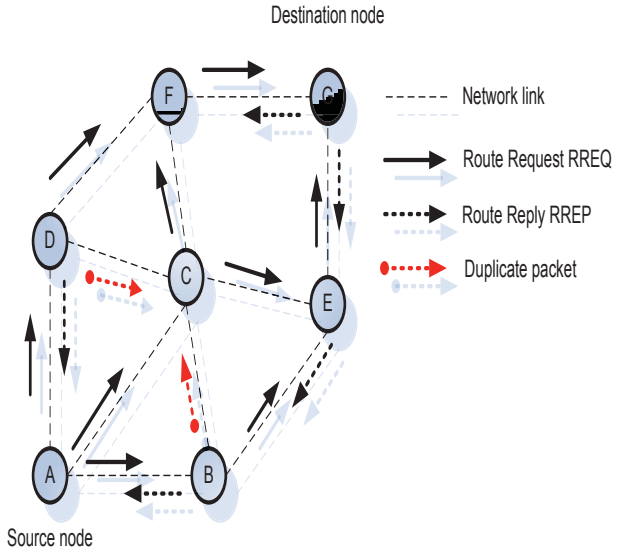


Figure 2: Route Discovery in DSR

- Route Maintenance

The route maintenance phase is carried out whenever there is a broken link between two nodes. A broken link can be detected by a node by either passively monitoring in promiscuous mode or actively monitoring the link. As shown in Figure 3, when a link break (F-G) happens, a route error packet (RERR) is sent by the intermediate node back to the originating node. The source node re-initiates the route discovery procedure to find a new route to the destination. It also removes any route entries it may have in its cache to that destination node.

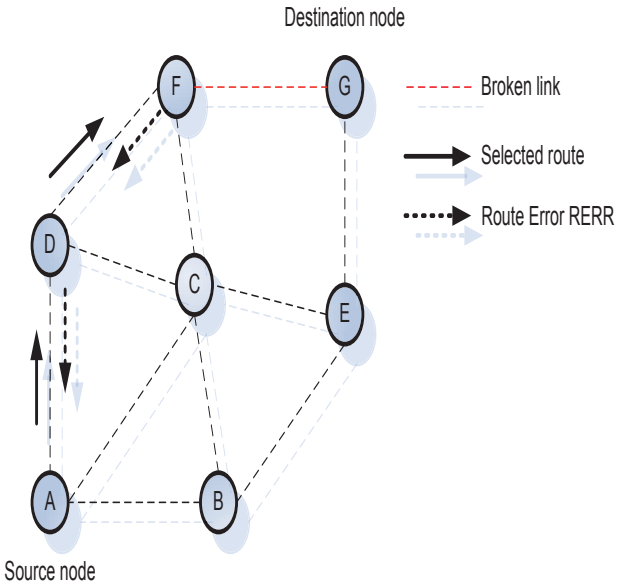


Figure 3: Route Maintenance in DSR

DSR benefits from source routing since the intermediate nodes do not need to maintain up-to-date routing information in order to route the packets that

they receive. There is also no need for any periodic routing advertisement messages. However, as size of the network increases, the routing overhead increases since each packet has to carry the entire route to the destination along with it. The use of route caches is a good mechanism to reduce the propagation delay but overuse of the cache may result in poor performance [7]. Another issue of DSR is that whenever there is a link break, the RERR packet propagates to the original source, which in turn initiates a new route discovery process. The link is not repaired locally. Several optimizations to DSR have been proposed, such as non- propagating route requests (when sending RREQ, nodes set the hop limit to one preventing them from re-broadcasting), gratuitous route replies (when a node overhears a packet with its own address listed in the header, it sends a RREP to the originating node bypassing the preceding hops), etc.

IV. AdHoc On Demand Distance Vector (AODV) Routing Protocol

To find routes, the AODV routing protocol [9] uses a reactive approach and to identify the most recent path it uses a proactive approach. That is, it uses the route discovery process similar to DSR to find routes and to compute fresh routes it uses destination sequence numbers. The two phases of the AODV routing protocol are described below.

- Route Discovery

In this phase, RREQ packets are transmitted by the source node in a way similar to DSR. The components of the RREQ packet include fields such as the source identifier (SId), the destination identifier (DId), the source sequence number (SSeq), the destination sequence number (DSeq), the broadcast identifier (BId), and TTL. When a RREQ packet is received by an intermediate node, it could either forward the RREQ packet or prepare a Route Reply (RREP) packet if there is an available valid route to the destination in its cache. To verify if a particular RREQ has already been received to avoid duplicates, the (SId, BId) pair is used. While transmitting a RREQ packet, every intermediate node enters the previous node's address and its BId. A timer associated with every entry is also maintained by the node in an attempt to delete a RREQ packet in case the reply has not been received before it expires.

When a node receives a RREP packet, the information of the previous node is also stored in it in order to forward the packet to it as the next hop of the destination. This plays a role of a "forward pointer" to the destination node. By doing it, each node contains only the next hop information; whereas in the source routing, all the intermediate nodes on the route towards the destination are stored.

Figure 4 depicts an example of route discovery mechanism in AODV. Suppose that node A wishes to forward a data packet to node G but it has not an available route in its cache. It then initiates a route discovery process by broadcasting a RREQ packet to all its neighboring nodes (B, C and D).

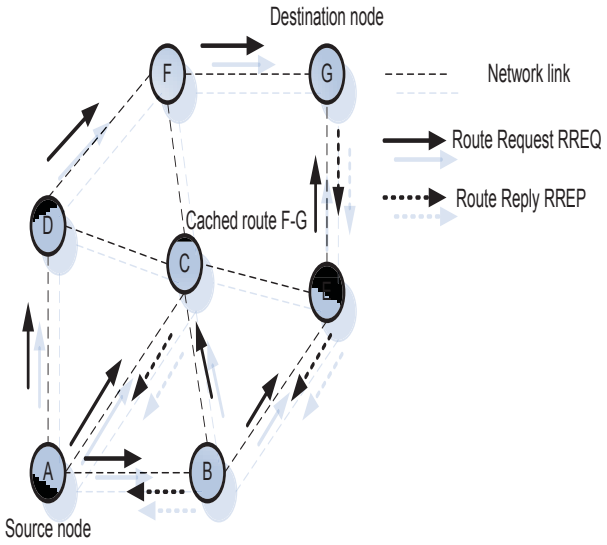


Figure 4: Route discovery in AODV

When RREQ packet reaches to nodes B, C and D, these nodes immediately search their respective route caches for an existing route. In the case where no route is available, they forward the RREQ to their neighbors; otherwise a comparison is made between the destination sequence number (DSeq) in the RREQ packet and the DSeq in its corresponding entry in the route cache. It replies to the source node with a RREP packet consisting of the route to the destination in the case the DSeq in the RREQ packet is greater.

- Route Maintenance

The way that the route maintenance mechanism works is described below. Whenever a node finds out a link break (via link layer acknowledgements or HELLO messages [9]), it broadcasts an RERR packet (in a way similar to DSR) to notify the source and the end nodes. This process is illustrated in Figure 5 If the link between nodes C and F breaks on the path A-C-F-G, RERR packets will be sent by both F and C to notify the source and the destination nodes.

The main advantage of AODV is the avoidance of source routing to reduce the routing overload in a large network. Another good feature of AODV is its application of expanding-ring-search to control the flood of RREQ packets and search for routes to unknown destinations [10]. In addition, it also supplies destination sequence numbers, allowing the nodes to have more up-to-date routes. However, some notes have to be taken into consideration when using AODV. Firstly, it requires bidirectional links and periodic link layer acknowledgements to detect broken links. Secondly, unlike DSR, it needs to maintain routing tables for route maintenance unlike DSR.

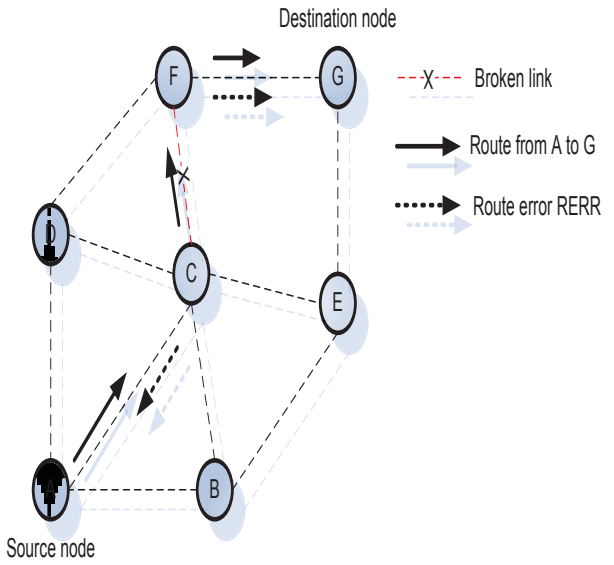


Figure 5: Route Maintenance in AODV

V. Attacks On Existing Protocols

In general, the attacks on routing protocols can generally be classified as routing disruption attacks [14][15] and resource consumption attacks [14][15]. In routing disruption attacks, the attacker tries to disrupt the routing mechanism by routing packets in wrong paths; in resource consumption attacks, some non-cooperative or selfish nodes may try to inject false packets in order to consume network bandwidth. Both of these attacks are examples of Denial of Service (DoS) attacks. Figure 6 depicts a broader classification of the possible attacks in MANETs.

- Attacks using Modification

In this type of attacks, some of the protocol fields of the messages passed among the nodes are modified, thereby resulting in traffic subversion, redirection or Denial of Service (DoS) attacks. The following sections discuss some of these attacks.

o ***Modification of Route Sequence Numbers :*** This attack is possible against the AODV protocol. The malicious node can change the sequence number in the route request packets or route reply packets in order to make the route fresh. In Figure 7, malicious node M receives a route request RREQ from node B that originates from node S and is destined for node X. M unicasts a RREP to B with a higher destination sequence number for X than the value last advertised by X. The node S accepts the RREP and then sends the data to X through M. When the legitimate RREP from X gets to S, if the destination number is less than the one advertised by M, then it will be discarded as a stale route. The situation will not be corrected until a valid RREP with higher sequence number than that of M gets to S.

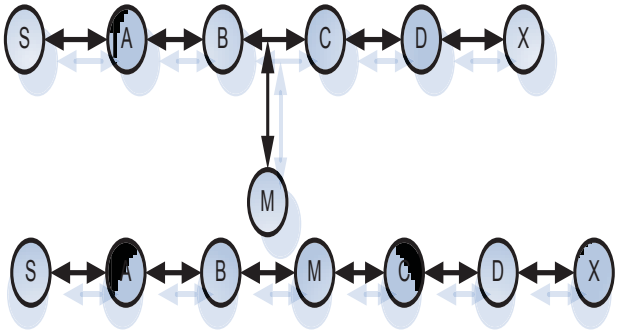


Figure 7: An example of route modification attack

o ***Modification of Hop Count :*** This type of attacks is possible against the AODV protocol in which a malicious node can increase the chance that they are included in a newly created route by resetting the hop count field of a RREQ packet to a lower number or even zero. Similar to route modification attack with sequence number, the hop count field in the routing packets is modified to attract data traffic.

o ***Modification of Source Route :*** This attack is possible against DSR which uses source routes and works as follows. In Figure 7, it is assumed that the shortest path exists from S to X. It is also assume that C and X cannot hear each other, that nodes B and C cannot hear each other, and that M is a malicious node attempting a denial-of-service attack. Suppose S sends a data packet to X with the source route S-A-B-C-D-X. If M intercepts this packet, it removes D from the list and forwards it to C. C will attempt to forward this packet to X which is not possible since C cannot hear X. Thus M has successfully launched a DoS attack on X.

- Attacks Using Impersonation

This type of attacks violates authenticity and confidentiality in a network. A malicious node can impersonate or spoof the address of another node in order to alter the vision of the network topology as perceived by another node. Such attacks can be described as follows in Figure 8

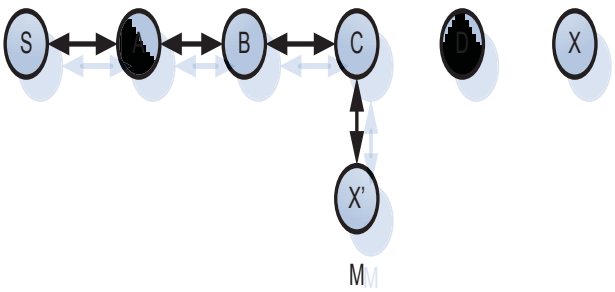


Figure 8: An example of impersonation attack

Node S wants to send data to node X and initiates a Route Discovery process. The malicious node M, closer to node S than node X, impersonates node X as X'. It sends a route reply (RREP) to node S. Without checking the authenticity of the RREP, node S accepts the route in the RREP and starts to send data to the malicious node. This type of attacks can cause a routing loop within the network.

- Special Attacks

In addition to the attacks described above, there are two other severe attacks which are possible against routing protocols such as AODV and DSR.

o ***Wormhole Attack*** : The wormhole attack [11] is a severe type of attacks in which two malicious nodes can forward packets through a private "*tunnel*" in the network as shown in Figure 9.

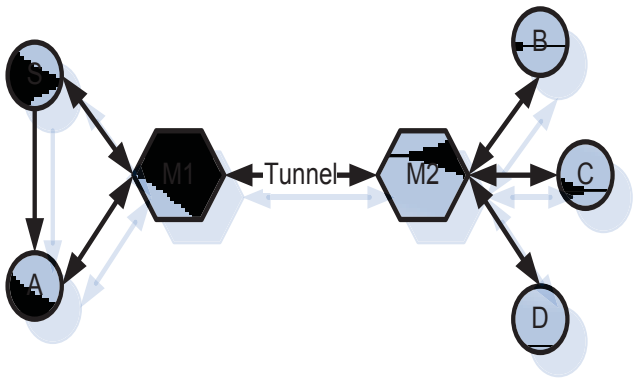


Figure 9: An example of wormhole attack

Here, M_1 and M_2 are two malicious nodes which link through a private connection. Every packet that M_1 receives from the network is forwarded through "wormhole" to node M_2, and vice versa. This attack disrupts routing protocols by short circuiting the normal flow of routing packets. Such a type of attack is difficult to detect in a network, and may severely damages the communication among the nodes. Such an attack can be prevented by using packet leashes [18], which authenticate the timing information in the packets to detect faked packets in the network.

o ***Black Hole Attack*** : A node advertises a zero metric for all destinations causing all nodes around it to route data packets towards it. The AODV protocol is vulnerable to such an attack.

VI. Experimental Results

The performance data of four routing protocols (DSR, ARIADNE, AODV and SAODV) are collected. A scenario is set up for data collection. This scenario is run 11 times with 11 different values of the mobility *pause time* ranging from 0 to 100 seconds. The data is collected according to two metrics

- Packet Delivery Fraction
- Normalized Routing Load

In general, the actual values of the performance metrics in a given scenario are affected by many factors, such as node speed, moving direction of the nodes, the destination of the traffic, data flow, congestion at a specific node, etc. It is therefore difficult to evaluate the performance of a protocol by directly comparing the acquired metrics from individual scenarios. In order to obtain representative values for the performance metrics, we decided to take the average values of multiple simulation runs. The average values of these 11 simulation runs are then calculated for the two metrics and used as a baseline to evaluate the performance of routing protocols in malicious environments.

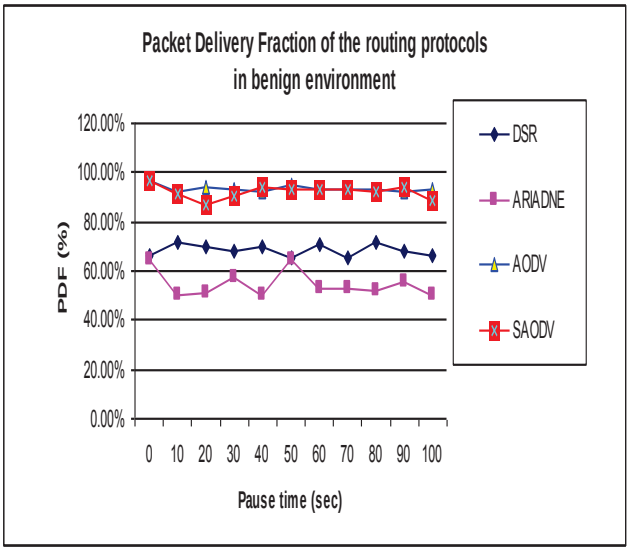


Figure 10: Packet Delivery Fraction vs. pause time values in benign environment

As shown in Figure 10 the percentage of packets delivered in AODV and SAODV is fairly close to each other, and both methods exhibit superior performance (~90% in general). The security features in SAODV lower the performance a little bit. Actually, the generation and verification of digital signatures depends on the power of the mobile nodes and causes a delay in routing packet processing. In the simulation environments, this delay depends on the simulation running machine and is not high enough to make the significant difference for the PDF metric. On the other hand, the packet delivery fraction in DSR and ARIADNE are 20-40% lower than that of AODV/SAODV across the board given different mobility pause times.

The major difference between AODV and DSR is caused by difference in their respective routing algorithms. It was reported by other researchers [5] [7] that, in high mobility and/or stressful data transmission scenarios, AODV outperforms DSR. The reason is that DSR heavily depends on the cached routes and lack any mechanism to expire stale routes. In the benign environment of our experiments, the default expiry timer

of cached route for DSR and ARIADNE is 300 seconds, while this number is 3 seconds for AODV and SAODV. In respect to the protocol design, these values are kept unchanged through all the simulation scenarios. Furthermore, DSR and ARIADNE store the complete path to the destination. Hence, if any node moves out of the communication range, the whole route becomes invalid. In MANETs, the nodes are mobile, so route change frequently occurs. Without being aware of most recent route changes, DSR may continue to send data packets along stale routes, leading to the increasing number of data packets being dropped.

The situation is even worse for ARIADNE, mainly because ARIADNE relies on the delayed key disclosure mechanism of TESLA when authenticating packets, including the RERR packets. When an intermediate node in ARIADNE notices a broken link, it sends a RERR message to the source node of the data packet. The source node, however, simply saves the RERR message, because it has not yet received from the intermediate node the key needed to authenticate the route error. The source node keeps sending the data until the second route error is triggered, and another RERR is received. Only then would the previous route error be authenticated, and the broken link not be used any more. This explains the worse performance of ARIADNE in comparison with DSR and other protocols.

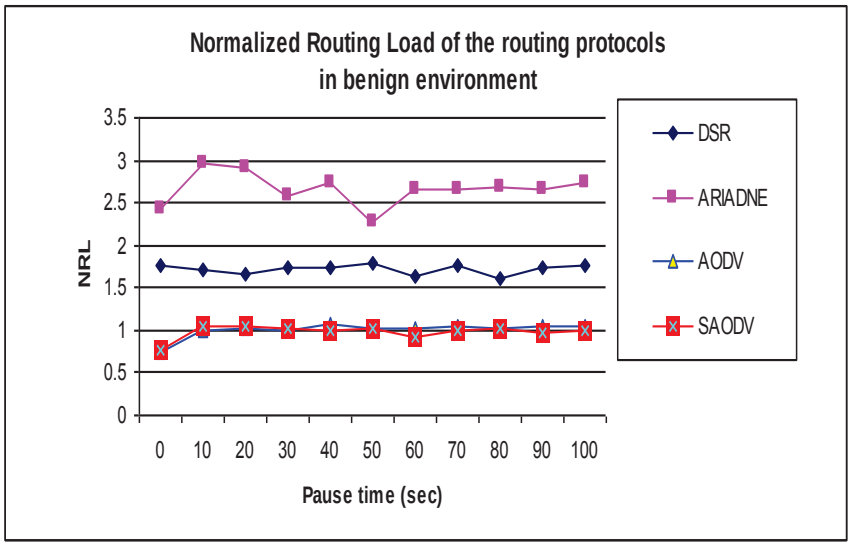


Figure 11: Normalized Routing Load vs. pause time values in benign environment

As shown in Figure 11, the NRL metric is, in general, inversely proportional to the PDF metric. A low PDF value corresponds to a high NRL value. This relationship between PDF and NRL is further illustrated in Table 1.1, which lists the average values of the two metrics over 11 simulation runs for each of the four protocols.

Pause Time (seconds)	Packet Delivery Fraction (%)	Normalized Routing Load
DSR	68.41%	1.72
ARIADNE	54.70%	2.58
AODV	93.45%	1.01
SAODV	92.00%	0.98

Table 1.1 The "baseline" metrics of the four protocols

VII. Conclusion

In this paper, I have implemented two secure routing protocols, ARIADNE and SAODV, based on their respective underlying protocols, DSR and AODV, in the OPNET simulation environment. I have also simulated four popular network attack models that exploit the weakness of the protocols. The attack models are used to make malicious wireless nodes and create various malicious environments, in which the performance of DSR, AODV, ARIADNE, and SAODV are evaluated. AODV and SAODV without pubic key verification are vulnerable to impersonation attacks. The impacts on the

two protocols are similar. The more the number of malicious nodes in the network is, the fewer the number of received data packets is. As shown by the experiments, SAODV is secure against impersonation attack only when there is a way to verify the public key of the route reply originator. In other words, a key management center is really necessary to make SAODV secure against impersonation attacks. This is still an outstanding issue of SAODV. The ultimate goal of a routing protocol is to efficiently deliver the network data to the destinations; therefore, two metrics, Packet Delivery Fraction (PDF) used to evaluate the protocols.

References Références Referencias

1. Yih-Chun Hu, Adrian Perrig, and David B. Johnson. "Ariadne: A secure On-Demand Routing Protocol for Ad hoc Networks". MobiCom 2002, September 23-28, 2002, Atlanta, Georgia, USA. http://www.ece.cmu.edu/~adrian/projects/secure routing/ariadne.pdf
2. Manel Guerrero Zapata. "Secure Ad hoc On-Demand Distance Vector Routing". ACM Mobile Computing and Communications Review (MC2R), 6(3):106--107, July 2002. http://lambda.cs.yale.edu/cs425/doc/zapata.pdf
3. Mishra Amitabh, Nadkarni Ketan M., and Ilyas Mohammad. "Chapter 30: Security in wireless ad-hoc networks, the handbook of Ad hoc wireless network", CRC PRESS Publisher, 2003.
4. Sadasivam Karthik, Vishal Changrani, T. Andrew Yang. "Scenario Based Performance Evaluation of Secure Routing in MANETs". http://sce.uhcl.edu/sadasivamk/MANETII05-draft.pdf
5. Josh Broch, David A. Maltz, David B. Johnson, Yih-Chun Hu and Jorjeta Jetcheva. "A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols". IEEE Journal on Selected Areas in Communication, 1999. http://monarch.cs.rice.edu/monarch-papers/mobicom98.pdf
6. Yuxia Lin, A. Hamed Mohsenian Rad, Vincent W.S. Wong, and Joo-Han Song. "Experimental Comparisons between SAODV and AODV Routing Protocols". In proceedings of the 1st ACM workshop on Wireless multimedia, 2005. http://www.ece.ubc.ca/~vincentw/C/LRWSc05.pdf
7. Samir R. Das, Charles E. Perkins, Elizabeth M. Royer. "Performance Comparison of Two On-demand Routing Protocols for Ad Hoc Networks". Proceedings IEEE Infocom page 3-12, March 2000. http://www.cs.ucsb.edu/~ebelding/txt/Perkins_PerfComp.pdf
8. D B. Johnson, D A. Maltz, and Y. Hu. "The dynamic source routing protocol for mobile ad hoc network,", Internet-Draft, April 2003. http://www.ietf.org/internet-drafts/draft-ietf-manet-dsr-09.txt
9. C.E. Perkins, E. Royer, and S.R. Das. "Ad hoc on demand distance vector (AODV) routing", Internet Draft, March 2000. http://www.ietf.org/internetdrafts/draft-ietf-manet-aodv-05.txt
10. C.Siva Ram Murthy and B. S. Manoj. "Ad hoc wireless networks: Architecture and Protocols". Prentice Hall Publishers, May 2004, ISBN 013147023X.
11. C.-K. Toh. "Ad Hoc Mobile Wireless Networks: Protocols and Systems". Prentice Hall publishers, December 2001, ISBN 0130078174.
12. David B. Johnson David A. Maltz Josh Broch. "DSR: The Dynamic Source Routing Protocol for Multi-Hop Wireless Ad Hoc Networks'.In Ad Hoc Networking, edited by Charles E. Perkins, Chapter 5, pp. 139-172, Addison-Wesley, 2001. http://www.cs.ust.hk/~qianzh/COMP680H/reading-list/johnson01.pdf
13. M. Marina and S. Das. "Performance or Route caching strategies in Dynamic Source Routing". Proceedings of the Int'l Workshop on Wireless Networks and Mobile Computing (WNMC) in conjunction with Int'l Conf. on Distributed Computing Systems (ICDCS), Washington, DC, USA, 2001. http://www.cs.utsa.edu/faculty/boppana/papers/phd-tdyer-dec02.pdf
14. Charles E. Perkins and Elizabeth M. Royer. "Ad hoc on demand Distance vector routing". Proceedings of the 2nd IEEE Workshop on Mobile Computing Systems and Applications, New Orleans, LA, February 1999, pp. 90-100. http://wisl.ece.cornell.edu/ECE794/Mar5/perkins-aodv.pdf
15. William Stallings. "Network Security essentials: Application and Standards", Pearson Education, Inc 2003, ISBN 0130351288.

Global Journals Inc. (US) Guidelines Handbook 2012

www.GlobalJournals.org

FELLOW OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (FARSC)

- 'FARSC' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'FARSC" can be added to name in the following manner. eg. **Dr. John E. Hall, Ph.D., FARSC or William Walldroff Ph. D., M.S., FARSC**
- Being FARSC is a respectful honor. It authenticates your research activities. After becoming FARSC, you can use 'FARSC' title as you use your degree in suffix of your name. This will definitely will enhance and add up your name. You can use it on your Career Counseling Materials/CV/Resume/Visiting Card/Name Plate etc.
- 60% Discount will be provided to FARSC members for publishing research papers in Global Journals Inc., if our Editorial Board and Peer Reviewers accept the paper. For the life time, if you are author/co-author of any paper bill sent to you will automatically be discounted one by 60%
- FARSC will be given a renowned, secure, free professional email address with 100 GB of space eg.johnhall@globaljournals.org. You will be facilitated with Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.
- FARSC member is eligible to become paid peer reviewer at Global Journals Inc. to earn up to 15% of realized author charges taken from author of respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account or to your PayPal account.
- Eg. If we had taken 420 USD from author, we can send 63 USD to your account.
- FARSC member can apply for free approval, grading and certification of some of their Educational and Institutional Degrees from Global Journals Inc. (US) and Open Association of Research, Society U.S.A.
- After you are FARSC. You can send us scanned copy of all of your documents. We will verify, grade and certify them within a month. It will be based on your academic records, quality of research papers published by you, and 50 more criteria. This is beneficial for your job interviews as recruiting organization need not just rely on you for authenticity and your unknown qualities, you would have authentic ranks of all of your documents. Our scale is unique worldwide.
- FARSC member can proceed to get benefits of free research podcasting in Global Research Radio with their research documents, slides and online movies.
- After your publication anywhere in the world, you can upload you research paper with your recorded voice or you can use our professional RJs to record your paper their voice. We can also stream your conference videos and display your slides online.
- FARSC will be eligible for free application of Standardization of their Researches by Open Scientific Standards. Standardization is next step and level after publishing in a journal. A team of research and professional will work with you to take your research to its next level, which is worldwide open standardization.

- FARSC is eligible to earn from their researches: While publishing his paper with Global Journals Inc. (US), FARSC can decide whether he/she would like to publish his/her research in closed manner. When readers will buy that individual research paper for reading, 80% of its earning by Global Journals Inc. (US) will be transferred to FARSC member's bank account after certain threshold balance. There is no time limit for collection. FARSC member can decide its price and we can help in decision.

MEMBER OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (MARSC)

- 'MARSC' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'MARSC" can be added to name in the following manner. eg. Dr. John E. Hall, Ph.D., MARSC or William Walldroff Ph. D., M.S., MARSC
- Being MARSC is a respectful honor. It authenticates your research activities. After becoming MARSC, you can use 'MARSC' title as you use your degree in suffix of your name. This will definitely will enhance and add up your name. You can use it on your Career Counseling Materials/CV/Resume/Visiting Card/Name Plate etc.
- 40% Discount will be provided to MARSC members for publishing research papers in Global Journals Inc., if our Editorial Board and Peer Reviewers accept the paper. For the life time, if you are author/co-author of any paper bill sent to you will automatically be discounted one by 60%
- MARSC will be given a renowned, secure, free professional email address with 30 GB of space eg.johnhall@globaljournals.org. You will be facilitated with Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.
- MARSC member is eligible to become paid peer reviewer at Global Journals Inc. to earn up to 10% of realized author charges taken from author of respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account or to your PayPal account.
- MARSC member can apply for free approval, grading and certification of some of their Educational and Institutional Degrees from Global Journals Inc. (US) and Open Association of Research, Society U.S.A.
- MARSC is eligible to earn from their researches: While publishing his paper with Global Journals Inc. (US), MARSC can decide whether he/she would like to publish his/her research in closed manner. When readers will buy that individual research paper for reading, 40% of its earning by Global Journals Inc. (US) will be transferred to MARSC member's bank account after certain threshold balance. There is no time limit for collection. MARSC member can decide its price and we can help in decision.

AUXILIARY MEMBERSHIPS

ANNUAL MEMBER

- Annual Member will be authorized to receive e-Journal GJMBR for one year (subscription for one year).
- The member will be allotted free 1 GB Web-space along with subDomain to contribute and participate in our activities.
- A professional email address will be allotted free 500 MB email space.

PAPER PUBLICATION

- The members can publish paper once. The paper will be sent to two-peer reviewer. The paper will be published after the acceptance of peer reviewers and Editorial Board.

Process of submission of Research Paper

The Area or field of specialization may or may not be of any category as mentioned in 'Scope of Journal' menu of the GlobalJournals.org website. There are 37 Research Journal categorized with Six parental Journals GJCST, GJMR, GJRE, GJMBR, GJSFR, GJHSS. For Authors should prefer the mentioned categories. There are three widely used systems UDC, DDC and LCC. The details are available as 'Knowledge Abstract' at Home page. The major advantage of this coding is that, the research work will be exposed to and shared with all over the world as we are being abstracted and indexed worldwide.

The paper should be in proper format. The format can be downloaded from first page of 'Author Guideline' Menu. The Author is expected to follow the general rules as mentioned in this menu. The paper should be written in MS-Word Format (*.DOC,*.DOCX).

The Author can submit the paper either online or offline. The authors should prefer online submission.Online Submission: There are three ways to submit your paper:

(A) (I) First, register yourself using top right corner of Home page then Login. If you are already registered, then login using your username and password.

(II) Choose corresponding Journal.

(III) Click 'Submit Manuscript'. Fill required information and Upload the paper.

(B) If you are using Internet Explorer, then Direct Submission through Homepage is also available.

(C) If these two are not convenient, and then email the paper directly to dean@globaljournals.org.

Offline Submission: Author can send the typed form of paper by Post. However, online submission should be preferred.

Preferred Author Guidelines

MANUSCRIPT STYLE INSTRUCTION (Must be strictly followed)

Page Size: 8.27" X 11'"

- Left Margin: 0.65
- Right Margin: 0.65
- Top Margin: 0.75
- Bottom Margin: 0.75
- Font type of all text should be Swis 721 Lt BT.
- Paper Title should be of Font Size 24 with one Column section.
- Author Name in Font Size of 11 with one column as of Title.
- Abstract Font size of 9 Bold, "Abstract" word in Italic Bold.
- Main Text: Font size 10 with justified two columns section
- Two Column with Equal Column with of 3.38 and Gaping of .2
- First Character must be three lines Drop capped.
- Paragraph before Spacing of 1 pt and After of 0 pt.
- Line Spacing of 1 pt
- Large Images must be in One Column
- Numbering of First Main Headings (Heading 1) must be in Roman Letters, Capital Letter, and Font Size of 10.
- Numbering of Second Main Headings (Heading 2) must be in Alphabets, Italic, and Font Size of 10.

You can use your own standard format also.
Author Guidelines:

1. General,

2. Ethical Guidelines,

3. Submission of Manuscripts,

4. Manuscript's Category,

5. Structure and Format of Manuscript,

6. After Acceptance.

1. GENERAL

Before submitting your research paper, one is advised to go through the details as mentioned in following heads. It will be beneficial, while peer reviewer justify your paper for publication.

Scope

The Global Journals Inc. (US) welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals Inc. (US) is parental platform for Global Journal of Computer Science and Technology, Researches in Engineering, Medical Research, Science Frontier Research, Human Social Science, Management, and Business organization. The choice of specific field can be done otherwise as following in Abstracting and Indexing Page on this Website. As the all Global

Journals Inc. (US) are being abstracted and indexed (in process) by most of the reputed organizations. Topics of only narrow interest will not be accepted unless they have wider potential or consequences.

2. ETHICAL GUIDELINES

Authors should follow the ethical guidelines as mentioned below for publication of research paper and research activities.

Papers are accepted on strict understanding that the material in whole or in part has not been, nor is being, considered for publication elsewhere. If the paper once accepted by Global Journals Inc. (US) and Editorial Board, will become the copyright of the Global Journals Inc. (US).

Authorship: The authors and coauthors should have active contribution to conception design, analysis and interpretation of findings. They should critically review the contents and drafting of the paper. All should approve the final version of the paper before submission

The Global Journals Inc. (US) follows the definition of authorship set up by the Global Academy of Research and Development. According to the Global Academy of R&D authorship, criteria must be based on:

1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.

2) Drafting the paper and revising it critically regarding important academic content.

3) Final approval of the version of the paper to be published.

All authors should have been credited according to their appropriate contribution in research activity and preparing paper. Contributors who do not match the criteria as authors may be mentioned under Acknowledgement.

Acknowledgements: Contributors to the research other than authors credited should be mentioned under acknowledgement. The specifications of the source of funding for the research if appropriate can be included. Suppliers of resources may be mentioned along with address.

Appeal of Decision: The Editorial Board's decision on publication of the paper is final and cannot be appealed elsewhere.

Permissions: It is the author's responsibility to have prior permission if all or parts of earlier published illustrations are used in this paper.

Please mention proper reference and appropriate acknowledgements wherever expected.

If all or parts of previously published illustrations are used, permission must be taken from the copyright holder concerned. It is the author's responsibility to take these in writing.

Approval for reproduction/modification of any information (including figures and tables) published elsewhere must be obtained by the authors/copyright holders before submission of the manuscript. Contributors (Authors) are responsible for any copyright fee involved.

3. SUBMISSION OF MANUSCRIPTS

Manuscripts should be uploaded via this online submission page. The online submission is most efficient method for submission of papers, as it enables rapid distribution of manuscripts and consequently speeds up the review procedure. It also enables authors to know the status of their own manuscripts by emailing us. Complete instructions for submitting a paper is available below.

Manuscript submission is a systematic procedure and little preparation is required beyond having all parts of your manuscript in a given format and a computer with an Internet connection and a Web browser. Full help and instructions are provided on-screen. As an author, you will be prompted for login and manuscript details as Field of Paper and then to upload your manuscript file(s) according to the instructions.

To avoid postal delays, all transaction is preferred by e-mail. A finished manuscript submission is confirmed by e-mail immediately and your paper enters the editorial process with no postal delays. When a conclusion is made about the publication of your paper by our Editorial Board, revisions can be submitted online with the same procedure, with an occasion to view and respond to all comments.

Complete support for both authors and co-author is provided.

4. MANUSCRIPT'S CATEGORY

Based on potential and nature, the manuscript can be categorized under the following heads:

Original research paper: Such papers are reports of high-level significant original research work.

Review papers: These are concise, significant but helpful and decisive topics for young researchers.

Research articles: These are handled with small investigation and applications

Research letters: The letters are small and concise comments on previously published matters.

5.STRUCTURE AND FORMAT OF MANUSCRIPT

The recommended size of original research paper is less than seven thousand words, review papers fewer than seven thousands words also.Preparation of research paper or how to write research paper, are major hurdle, while writing manuscript. The research articles and research letters should be fewer than three thousand words, the structure original research paper; sometime review paper should be as follows:

Papers: These are reports of significant research (typically less than 7000 words equivalent, including tables, figures, references), and comprise:

(a)Title should be relevant and commensurate with the theme of the paper.

(b) A brief Summary, "Abstract" (less than 150 words) containing the major results and conclusions.

(c) Up to ten keywords, that precisely identifies the paper's subject, purpose, and focus.

(d) An Introduction, giving necessary background excluding subheadings; objectives must be clearly declared.

(e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.

(f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;

(g) Discussion should cover the implications and consequences, not just recapitulating the results; conclusions should be summarizing.

(h) Brief Acknowledgements.

(i) References in the proper form.

Authors should very cautiously consider the preparation of papers to ensure that they communicate efficiently. Papers are much more likely to be accepted, if they are cautiously designed and laid out, contain few or no errors, are summarizing, and be conventional to the approach and instructions. They will in addition, be published with much less delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and to make suggestions to improve briefness.

It is vital, that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

Format

Language: The language of publication is UK English. Authors, for whom English is a second language, must have their manuscript efficiently edited by an English-speaking person before submission to make sure that, the English is of high excellence. It is preferable, that manuscripts should be professionally edited.

Standard Usage, Abbreviations, and Units: Spelling and hyphenation should be conventional to The Concise Oxford English Dictionary. Statistics and measurements should at all times be given in figures, e.g. 16 min, except for when the number begins a sentence. When the number does not refer to a unit of measurement it should be spelt in full unless, it is 160 or greater.

Abbreviations supposed to be used carefully. The abbreviated name or expression is supposed to be cited in full at first usage, followed by the conventional abbreviation in parentheses.

Metric SI units are supposed to generally be used excluding where they conflict with current practice or are confusing. For illustration, 1.4 l rather than 1.4 × 10-3 m3, or 4 mm somewhat than 4 × 10-3 m. Chemical formula and solutions must identify the form used, e.g. anhydrous or hydrated, and the concentration must be in clearly defined units. Common species names should be followed by underlines at the first mention. For following use the generic name should be constricted to a single letter, if it is clear.

Structure

All manuscripts submitted to Global Journals Inc. (US), ought to include:

Title: The title page must carry an instructive title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) wherever the work was carried out. The full postal address in addition with the e-mail address of related author must be given. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining and indexing.

Abstract, used in Original Papers and Reviews:

Optimizing Abstract for Search Engines

Many researchers searching for information online will use search engines such as Google, Yahoo or similar. By optimizing your paper for search engines, you will amplify the chance of someone finding it. This in turn will make it more likely to be viewed and/or cited in a further work. Global Journals Inc. (US) have compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Key Words

A major linchpin in research work for the writing research paper is the keyword search, which one will employ to find both library and Internet resources.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy and planning a list of possible keywords and phrases to try.

Search engines for most searches, use Boolean searching, which is somewhat different from Internet searches. The Boolean search uses "operators," words (and, or, not, and near) that enable you to expand or narrow your affords. Tips for research paper while preparing research paper are very helpful guideline of research paper.

Choice of key words is first tool of tips to write research paper. Research paper writing is an art.A few tips for deciding as strategically as possible about keyword search:

- One should start brainstorming lists of possible keywords before even begin searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in research paper?" Then consider synonyms for the important words.
- It may take the discovery of only one relevant paper to let steer in the right keyword direction because in most databases, the keywords under which a research paper is abstracted are listed with the paper.
- One should avoid outdated words.

Keywords are the key that opens a door to research work sources. Keyword searching is an art in which researcher's skills are bound to improve with experience and time.

Numerical Methods: Numerical methods used should be clear and, where appropriate, supported by references.

Acknowledgements: Please make these as concise as possible.

References

References follow the Harvard scheme of referencing. References in the text should cite the authors' names followed by the time of their publication, unless there are three or more authors when simply the first author's name is quoted followed by et al. unpublished work has to only be cited where necessary, and only in the text. Copies of references in press in other journals have to be supplied with submitted typescripts. It is necessary that all citations and references be carefully checked before submission, as mistakes or omissions will cause delays.

References to information on the World Wide Web can be given, but only if the information is available without charge to readers on an official site. Wikipedia and Similar websites are not allowed where anyone can change the information. Authors will be asked to make available electronic copies of the cited information for inclusion on the Global Journals Inc. (US) homepage at the judgment of the Editorial Board.

The Editorial Board and Global Journals Inc. (US) recommend that, citation of online-published papers and other material should be done via a DOI (digital object identifier). If an author cites anything, which does not have a DOI, they run the risk of the cited material not being noticeable.

The Editorial Board and Global Journals Inc. (US) recommend the use of a tool such as Reference Manager for reference management and formatting.

Tables, Figures and Figure Legends

Tables: Tables should be few in number, cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g. Table 4, a self-explanatory caption and be on a separate sheet. Vertical lines should not be used.

Figures: Figures are supposed to be submitted as separate files. Always take in a citation in the text for each figure using Arabic numbers, e.g. Fig. 4. Artwork must be submitted online in electronic form by e-mailing them.

Preparation of Electronic Figures for Publication

Even though low quality images are sufficient for review purposes, print publication requires high quality images to prevent the final product being blurred or fuzzy. Submit (or e-mail) EPS (line art) or TIFF (halftone/photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Do not use pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings) in relation to the imitation size. Please give the data for figures in black and white or submit a Color Work Agreement Form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution (at final image size) ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs) : >350 dpi; figures containing both halftone and line images: >650 dpi.

Color Charges: It is the rule of the Global Journals Inc. (US) for authors to pay the full cost for the reproduction of their color artwork. Hence, please note that, if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a color work agreement form before your paper can be published.

Figure Legends: Self-explanatory legends of all figures should be incorporated separately under the heading 'Legends to Figures'. In the full-text online edition of the journal, figure legends may possibly be truncated in abbreviated links to the full screen version. Therefore, the first 100 characters of any legend should notify the reader, about the key aspects of the figure.

6. AFTER ACCEPTANCE

Upon approval of a paper for publication, the manuscript will be forwarded to the dean, who is responsible for the publication of the Global Journals Inc. (US).

6.1 Proof Corrections

The corresponding author will receive an e-mail alert containing a link to a website or will be attached. A working e-mail address must therefore be provided for the related author.

Acrobat Reader will be required in order to read this file. This software can be downloaded

(Free of charge) from the following website:

www.adobe.com/products/acrobat/readstep2.html. This will facilitate the file to be opened, read on screen, and printed out in order for any corrections to be added. Further instructions will be sent with the proof.

Proofs must be returned to the dean at dean@globaljournals.org within three days of receipt.

As changes to proofs are costly, we inquire that you only correct typesetting errors. All illustrations are retained by the publisher. Please note that the authors are responsible for all statements made in their work, including changes made by the copy editor.

6.2 Early View of Global Journals Inc. (US) (Publication Prior to Print)

The Global Journals Inc. (US) are enclosed by our publishing's Early View service. Early View articles are complete full-text articles sent in advance of their publication. Early View articles are absolute and final. They have been completely reviewed, revised and edited for publication, and the authors' final corrections have been incorporated. Because they are in final form, no changes can be made after sending them. The nature of Early View articles means that they do not yet have volume, issue or page numbers, so Early View articles cannot be cited in the conventional way.

6.3 Author Services

Online production tracking is available for your article through Author Services. Author Services enables authors to track their article - once it has been accepted - through the production process to publication online and in print. Authors can check the status of their articles online and choose to receive automated e-mails at key stages of production. The authors will receive an e-mail with a unique link that enables them to register and have their article automatically added to the system. Please ensure that a complete e-mail address is provided when submitting the manuscript.

6.4 Author Material Archive Policy

Please note that if not specifically requested, publisher will dispose off hardcopy & electronic information submitted, after the two months of publication. If you require the return of any information submitted, please inform the Editorial Board or dean as soon as possible.

6.5 Offprint and Extra Copies

A PDF offprint of the online-published article will be provided free of charge to the related author, and may be distributed according to the Publisher's terms and conditions. Additional paper offprint may be ordered by emailing us at: editor@globaljournals.org .

the search? Will I be able to find all information in this field area? If the answer of these types of questions will be "Yes" then you can choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

2. Evaluators are human: First thing to remember that evaluators are also human being. They are not only meant for rejecting a paper. They are here to evaluate your paper. So, present your Best.

3. Think Like Evaluators: If you are in a confusion or getting demotivated that your paper will be accepted by evaluators or not, then think and try to evaluate your paper like an Evaluator. Try to understand that what an evaluator wants in your research paper and automatically you will have your answer.

4. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

5. Ask your Guides: If you are having any difficulty in your research, then do not hesitate to share your difficulty to your guide (if you have any). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work then ask the supervisor to help you with the alternative. He might also provide you the list of essential readings.

6. Use of computer is recommended: As you are doing research in the field of Computer Science, then this point is quite obvious.

7. Use right software: Always use good quality software packages. If you are not capable to judge good software then you can lose quality of your paper unknowingly. There are various software programs available to help you, which you can get through Internet.

8. Use the Internet for help: An excellent start for your paper can be by using the Google. It is an excellent search engine, where you can have your doubts resolved. You may also read some answers for the frequent question how to write my research paper or find model research paper. From the internet library you can download books. If you have all required books make important reading selecting and analyzing the specified information. Then put together research paper sketch out.

9. Use and get big pictures: Always use encyclopedias, Wikipedia to get pictures so that you can go into the depth.

10. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right! It is a good habit, which helps to not to lose your continuity. You should always use bookmarks while searching on Internet also, which will make your search easier.

11. Revise what you wrote: When you write anything, always read it, summarize it and then finalize it.

12. Make all efforts: Make all efforts to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in introduction, that what is the need of a particular research paper. Polish your work by good skill of writing and always give an evaluator, what he wants.

13. Have backups: When you are going to do any important thing like making research paper, you should always have backup copies of it either in your computer or in paper. This will help you to not to lose any of your important.

14. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several and unnecessary diagrams will degrade the quality of your paper by creating "hotchpotch." So always, try to make and include those diagrams, which are made by your own to improve readability and understandability of your paper.

15. Use of direct quotes: When you do research relevant to literature, history or current affairs then use of quotes become essential but if study is relevant to science then use of quotes is not preferable.

16. Use proper verb tense: Use proper verb tenses in your paper. Use past tense, to present those events that happened. Use present tense to indicate events that are going on. Use future tense to indicate future happening events. Use of improper and wrong tenses will confuse the evaluator. Avoid the sentences that are incomplete.

17. Never use online paper: If you are getting any paper on Internet, then never use it as your research paper because it might be possible that evaluator has already seen it or maybe it is outdated version.

18. **Pick a good study spot:** To do your research studies always try to pick a spot, which is quiet. Every spot is not for studies. Spot that suits you choose it and proceed further.

19. Know what you know: Always try to know, what you know by making objectives. Else, you will be confused and cannot achieve your target.

20. Use good quality grammar: Always use a good quality grammar and use words that will throw positive impact on evaluator. Use of good quality grammar does not mean to use tough words, that for each word the evaluator has to go through dictionary. Do not start sentence with a conjunction. Do not fragment sentences. Eliminate one-word sentences. Ignore passive voice. Do not ever use a big word when a diminutive one would suffice. Verbs have to be in agreement with their subjects. Prepositions are not expressions to finish sentences with. It is incorrect to ever divide an infinitive. Avoid clichés like the disease. Also, always shun irritating alliteration. Use language that is simple and straight forward. put together a neat summary.

21. Arrangement of information: Each section of the main body should start with an opening sentence and there should be a changeover at the end of the section. Give only valid and powerful arguments to your topic. You may also maintain your arguments with records.

22. Never start in last minute: Always start at right time and give enough time to research work. Leaving everything to the last minute will degrade your paper and spoil your work.

23. Multitasking in research is not good: Doing several things at the same time proves bad habit in case of research activity. Research is an area, where everything has a particular time slot. Divide your research work in parts and do particular part in particular time slot.

24. Never copy others' work: Never copy others' work and give it your name because if evaluator has seen it anywhere you will be in trouble.

25. Take proper rest and food: No matter how many hours you spend for your research activity, if you are not taking care of your health then all your efforts will be in vain. For a quality research, study is must, and this can be done by taking proper rest and food.

26. Go for seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

27. Refresh your mind after intervals: Try to give rest to your mind by listening to soft music or by sleeping in intervals. This will also improve your memory.

28. Make colleagues: Always try to make colleagues. No matter how sharper or intelligent you are, if you make colleagues you can have several ideas, which will be helpful for your research.

29. **Think technically:** Always think technically. If anything happens, then search its reasons, its benefits, and demerits.

30. Think and then print: When you will go to print your paper, notice that tables are not be split, headings are not detached from their descriptions, and page sequence is maintained.

31. Adding unnecessary information: Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be

sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

32. Never oversimplify everything: To add material in your research paper, never go for oversimplification. This will definitely irritate the evaluator. Be more or less specific. Also too, by no means, ever use rhythmic redundancies. Contractions aren't essential and shouldn't be there used. Comparisons are as terrible as clichés. Give up ampersands and abbreviations, and so on. Remove commas, that are, not necessary. Parenthetical words however should be together with this in commas. Understatement is all the time the complete best way to put onward earth-shaking thoughts. Give a detailed literary review.

33. Report concluded results: Use concluded results. From raw data, filter the results and then conclude your studies based on measurements and observations taken. Significant figures and appropriate number of decimal places should be used. Parenthetical remarks are prohibitive. Proofread carefully at final stage. In the end give outline to your arguments. Spot out perspectives of further study of this subject. Justify your conclusion by at the bottom of them with sufficient justifications and examples.

34. After conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print to the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects in your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form, which is presented in the guidelines using the template.
- Please note the criterion for grading the final paper by peer-reviewers.

Final Points:

A purpose of organizing a research paper is to let people to interpret your effort selectively. The journal requires the following sections, submitted in the order listed, each section to start on a new page.

The introduction will be compiled from reference matter and will reflect the design processes or outline of basis that direct you to make study. As you will carry out the process of study, the method and process section will be constructed as like that. The result segment will show related statistics in nearly sequential order and will direct the reviewers next to the similar intellectual paths throughout the data that you took to carry out your study. The discussion section will provide understanding of the data and projections as to the implication of the results. The use of good quality references all through the paper will give the effort trustworthiness by representing an alertness of prior workings.

Writing a research paper is not an easy job no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record keeping are the only means to make straightforward the progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

- Insertion a title at the foot of a page with the subsequent text on the next page

- Separating a table/chart or figure - impound each figure/table to a single page
- Submitting a manuscript with pages out of sequence

In every sections of your document

· Use standard writing style including articles ("a", "the," etc.)

· Keep on paying attention on the research topic of the paper

· Use paragraphs to split each significant point (excluding for the abstract)

· Align the primary line of each section

· Present your points in sound order

· Use present tense to report well accepted

· Use past tense to describe specific results

· Shun familiar wording, don't address the reviewer directly, and don't use slang, slang language, or superlatives

· Shun use of extra pictures - include only those figures essential to presenting results

Title Page:

Choose a revealing title. It should be short. It should not have non-standard acronyms or abbreviations. It should not exceed two printed lines. It should include the name(s) and address (es) of all authors.

Abstract:

The summary should be two hundred words or less. It should briefly and clearly explain the key findings reported in the manuscript--must have precise statistics. It should not have abnormal acronyms or abbreviations. It should be logical in itself. Shun citing references at this point.

An abstract is a brief distinct paragraph summary of finished work or work in development. In a minute or less a reviewer can be taught the foundation behind the study, common approach to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Yet, use comprehensive sentences and do not let go readability for briefness. You can maintain it succinct by phrasing sentences so that they provide more than lone rationale. The author can at this moment go straight to

shortening the outcome. Sum up the study, with the subsequent elements in any summary. Try to maintain the initial two items to no more than one ruling each.

- Reason of the study - theory, overall issue, purpose
- Fundamental goal
- To the point depiction of the research
- Consequences, including definite statistics - if the consequences are quantitative in nature, account quantitative data; results of any numerical analysis should be reported
- Significant conclusions or questions that track from the research(es)

Approach:

- Single section, and succinct
- As a outline of job done, it is always written in past tense
- A conceptual should situate on its own, and not submit to any other part of the paper such as a form or table
- Center on shortening results - bound background information to a verdict or two, if completely necessary
- What you account in an conceptual must be regular with what you reported in the manuscript
- Exact spelling, clearness of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else

Introduction:

The **Introduction** should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable to comprehend and calculate the purpose of your study without having to submit to other works. The basis for the study should be offered. Give most important references but shun difficult to make a comprehensive appraisal of the topic. In the introduction, describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will have no attention in your result. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here. Following approach can create a valuable beginning:

- Explain the value (significance) of the study
- Shield the model - why did you employ this particular system or method? What is its compensation? You strength remark on its appropriateness from a abstract point of vision as well as point out sensible reasons for using it.
- Present a justification. Status your particular theory (es) or aim(s), and describe the logic that led you to choose them.
- Very for a short time explain the tentative propose and how it skilled the declared objectives.

Approach:

- Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done.
- Sort out your thoughts; manufacture one key point with every section. If you make the four points listed above, you will need a least of four paragraphs.
- Present surroundings information only as desirable in order hold up a situation. The reviewer does not desire to read the whole thing you know about a topic.
- Shape the theory/purpose specifically - do not take a broad view.
- As always, give awareness to spelling, simplicity and correctness of sentences and phrases.

Procedures (Methods and Materials):

This part is supposed to be the easiest to carve if you have good skills. A sound written Procedures segment allows a capable scientist to replacement your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt for the least amount of information that would permit another capable scientist to spare your outcome but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section. When a technique is used that has been well described in another object, mention the specific item describing a way but draw the basic

principle while stating the situation. The purpose is to text all particular resources and broad procedures, so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step by step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

- Explain materials individually only if the study is so complex that it saves liberty this way.
- Embrace particular materials, and any tools or provisions that are not frequently found in laboratories.
- Do not take in frequently found.
- If use of a definite type of tools.
- Materials may be reported in a part section or else they may be recognized along with your measures.

Methods:

- Report the method (not particulars of each process that engaged the same methodology)
- Describe the method entirely
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures
- Simplify - details how procedures were completed not how they were exclusively performed on a particular day.
- If well known procedures were used, account the procedure by name, possibly with reference, and that's all.

Approach:

- It is embarrassed or not possible to use vigorous voice when documenting methods with no using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result when script up the methods most authors use third person passive voice.
- Use standard style in this and in every other part of the paper - avoid familiar lists, and use full sentences.

What to keep away from

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings - save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part a entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Carry on to be to the point, by means of statistics and tables, if suitable, to present consequences most efficiently.You must obviously differentiate material that would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matter should not be submitted at all except requested by the instructor.

Content

- Sum up your conclusion in text and demonstrate them, if suitable, with figures and tables.
- In manuscript, explain each of your consequences, point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation an exacting study.
- Explain results of control experiments and comprise remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or in manuscript form.

What to stay away from

- Do not discuss or infer your outcome, report surroundings information, or try to explain anything.
- Not at all, take in raw data or intermediate calculations in a research manuscript.

- Do not present the similar data more than once.
- Manuscript should complement any figures or tables, not duplicate the identical information.
- Never confuse figures with tables - there is a difference.

Approach

- As forever, use past tense when you submit to your results, and put the whole thing in a reasonable order.
- Put figures and tables, appropriately numbered, in order at the end of the report
- If you desire, you may place your figures and tables properly within the text of your results part.

Figures and tables

- If you put figures and tables at the end of the details, make certain that they are visibly distinguished from any attach appendix materials, such as raw facts
- Despite of position, each figure must be numbered one after the other and complete with subtitle
- In spite of position, each table must be titled, numbered one after the other and complete with heading
- All figure and table must be adequately complete that it could situate on its own, divide from text

Discussion:

The Discussion is expected the trickiest segment to write and describe. A lot of papers submitted for journal are discarded based on problems with the Discussion. There is no head of state for how long a argument should be. Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implication of the study. The purpose here is to offer an understanding of your results and hold up for all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of result should be visibly described. Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved with prospect, and let it drop at that.

- Make a decision if each premise is supported, discarded, or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."
- Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work
- You may propose future guidelines, such as how the experiment might be personalized to accomplish a new idea.
- Give details all of your remarks as much as possible, focus on mechanisms.
- Make a decision if the tentative design sufficiently addressed the theory, and whether or not it was correctly restricted.
- Try to present substitute explanations if sensible alternatives be present.
- One research will not counter an overall question, so maintain the large picture in mind, where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

- When you refer to information, differentiate data generated by your own studies from available information
- Submit to work done by specific persons (including you) in past tense.
- Submit to generally acknowledged facts and main beliefs in present tense.

Administration Rules Listed Before Submitting Your Research Paper to Global Journals Inc. (US)

Please carefully note down following rules and regulation before submitting your Research Paper to Global Journals Inc. (US):

Segment Draft and Final Research Paper: You have to strictly follow the template of research paper. If it is not done your paper may get rejected.

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.

- Do not give permission to anyone else to "PROOFREAD" your manuscript.

- Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.

Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

INDEX

Q

R

S

T

U

V

W

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350