5-3-2012

International Joint Conference On Emerging Intelligent Sustainable Technologies (EISTCON-2012) VOLUME-3

Prof. Srikanta Patnaik Mentor

International Joint Conference

On

EMERGING INTELLIGENT SUSTAINABLE TECHNOLOGIES

(EISTCON-2012)

VOLUME-3

3rd & 4th May, 2012

Organised By:

Department Of Electrical & Electronics Engineering

Dayananda Sagar College of Engineering

Bangalore-560078, India

&

Interscience Research Network (IRNet)

Bhubaneswar, India

Mail To: irnet.chapter@gmail.com

Venue:

Dayananda Sagar College of Engineering

Shavige Malleswara Hills, Kumaraswamy Layout, Bangalore-560078

About: EISTCON-2012

There is no second thought that the academic and industrial scenario of 21st century is a sphere of knowledge driven by cutting edge technologies. It appears like Sky is the limit for the aspiring generation. The scope of today's amazing technology enables us to diffuse the ideas and transform the living conditions. The prospering dream to master over economic resources available the mankind has developed and categorised a concrete knowledge system that gives solutions and support to socio-economic and techno structural problems of our loving and enriching society. Some disciplines have their own theoretical foundations and some are derived. The growing interdependence of differentiating knowledge from several domains is leading an interdisciplinary approach or a holistic approach in analysing technical or social problems. . The thrust is to propagate new research and learning from experts, practitioners and avid scholars and improving the field of study for future development. The broad framework in which the conference is manifested is to augment and cater to the purpose of those silent laboratories and their relentless insistence for coming out with flying colours. The conference also honours to the most valued and contributing articles through its **"Young Investigator Award"** scheme. The areas covered under the conference are as follows:

- Stream-1: Electrical Engineering
- Stream-2: Electronics & Communication Engineering
- Stream-3: Computer Science Engg. & Information Technology

Stream-1: Electrical Engineering

The conference invites papers from all areas under the broad spectrum of power system planning, coordination, marketing, safety, stability, security, optimality, economy, reliability and their applications. The areas are as follows:

- Power system modeling, simulation and analysis
- Power system stability, dynamics and control
- Power generation, transmission and distribution
- FACTS & HVDC
- Smart Grid Technology
- Parallel Processing in Power Systems
- High voltage engineering and insulation coordination
- Distributed generation systems, micro grid operation and wind power systems
- Power Quality and demand side management
- Power system protection and digital relaying
- Energy management policy planning and decision making
- Network restructuring and marketing strategy
- Power system optimization
- State estimation and security analysis
- Reliability analysis
- Modeling uncertainties and contingencies
- Fault detection and diagnosis
- Deregulation of electricity market

- Distribution automation
- Machines, power electronics and drives
- Computer applications in power systems
- Soft computing applications in power systems
- Power system communication
- Renewable Energy Sources & Management
- Hybrid energy systems
- Power system instrumentation
- Energy conservation and management
- Fuel cells and batteries
- Nuclear power alternatives
- Remote sensing, telemetry, and signal processing
- Application of Nano-Technology
- Use of super conductors in power systems
- Energy efficient protocols, tools and gadgets
- Microprocessor/Microcontroller based applications
- Supervisory control and data acquisition

Stream-2: Electronics & Communication Engineering:

The conference covers all aspects of theory and design of circuits, systems and devices for electronics, signal processing, and communication, including:

- Signal and System theory, Digital signal processing
- Network theory and Circuit design
- Information theory, Communication theory and techniques, Modulation, Source and Channel coding
- Switching theory and techniques, Communication protocols
- Optical communications
- Microwave theory and techniques, Radar, Sonar
- Antennas, wave propagation
- Measurement and instrumentation; Circuit design, Simulation and CAD
- Signal and Image processing, Coding; Microwaves, Antennas and Radio propagation
- Optoelectronics; TV and Sound broadcasting; Telecommunication networks; Radio and Satellite communications; Radar, Sonar and navigation systems; Electromagnetic compatibility.

Stream-3: Computer Science Engineering. & Information Technology:

The conference aims to bring together developers, users, academicians and researchers in the information technology and computer science for sharing and exploring new areas of research and development and to discuss emerging issues faced by them.

- Algorithms Artificial Intelligence
- Automated Software Engineering
- Bioinformatics and Scientific Computing
- Biomedical Engineering
- Compilers and Interpreters
- E-commerce and E-governance
- Event Driven Programming
- Expert Systems
- High Performance Computing
- Human Computer Interaction

- Computational Intelligence
- Computer Animation
- Computer Architecture & VLSI
- Computer Architecture and Embedded Systems
- Computer Based Education
- Computer Games
- Computer Graphics & Virtual Reality
- Computer Graphics and Multimedia
- Computer Networks,Modeling & Data Communication
- Computer Security / Simulation/ Vision
- Computer-aided Design/Manufacturing
- Computing Ethics/ Practices & Applications
- Control Systems
- Data Communications / Compression/ Encryption.
- Data Mining
- Database Systems
- Digital Library
- Digital Signal and Image Processing
- Digital System and Logic Design
- Distributed and Parallel Processing
- Distributed Systems
- Information Retrieval Systems
- Internet and Web Applications
- Knowledge Data Engineering
- Mobile Computing
- Multimedia Applications
- Natural Language
- Processing Neural Networks
- Parallel and Distributed Computing
- Pattern Recognition
- Performance Evaluation
- Programming Languages
- Reconfigurable Computing Systems
- Robotics and Automation Security & Cryptography
- Software Engineering & CASE
- System Security
- Technology in Education Technology
- Management Theoretical Computer Science
- Ubiquitous Computing
- Wireless Sensor Networks
- Wireless Communication and Mobile Computing

COMMITTEE MEMBERS

Program Chair /Chief Mentor:

Dr. K. Shanmukha Sundar
Prof & HOD.
Electrical & Electronics Engineering Department,
Dayananda Sagar College of Engineering
Bangalore-560078, India
E-mail: bonniekhs@gmail.com

Prof. (Dr.) Srikanta Patnaik
President, IRNet India and Chairman IIMT
Intersceince Campus, Bhubaneswar
Email: srikantapatnaik@hotmail.com

Program Co-Chairs:

Dr.B. N. Sarkar
Additional Directro(Retd),CPRI,
Bangalore,India

Secretary IRNet:

Prof. Pradeep Kumar Mallick
Department of Computer Science
IIMT, Bhubaneswar
Email:pradeepmallick84@gmail.com
Mobile No: 08895885152

Conference Secretary:

Dr. Vasantha Kumar.S
EEE Department, DSCE,Bangalore

Conference Coordinator:
Suchana Mishra
EEE Department, DSCE, Bangalore

Program Coordinator:

Balaji Ramakrishna
EEE Department, DSCE,Bangalore

Madhura .S
EEE Department, DSCE,Bangalore

Technical Committee:

Dr. B. R. Lakshmikantha,
Principal, Dayananda Sagar Academy of Technology
Bangalore- 560082,India

Dr. B. N. Sarkar
Add Director (Retd),CPRI,Bangalore
Dr. Ravishankar Dixit
Prof & HOD, EEE Dept., BMSCE,Basavana Gudi
Bangalore

Dr. H. M. Ravikumar
Prof &HOD, EEE Dept.
NITTE Meenakshi Institute of Technology
Yelahanka, Bangalore - 560 064.

Dr. Pradeepkumar Dixit
MSRIT, Bangalore,India

Dr. K. L. Putta Buddhi
Professor,SJCE, Mysore

Smt. R. V. Parimala
Asst. Professor, DSCE,Bangalore

Smt. P. Usha
Asst. Professor, DSCE,Bangalore

Dr. B. K. Keshavan
Prof & HOD,EEE Dept.,
PESIT, Bangalore

Dr. Basavaraj Banakar
Vice Principal, Prof & HOD , GITAM University,Hyderabad

Post Conference Coordinator:
Miss. Ujjayinee Swain
IRNet , Bhubaneswar,India
Mob:08763722892

Head System & Utilities:
Prof. Sanjay Sharma
IIMT, Bhubaneswar
Mr. Sharath Kumar
EEE Department, DSCE,Bangalore

Publicity Committee:
Mr. Manjunatha Babu
EEE Department,DSCE,Bangalore
Mr. Rajaram
EEE Department,DSCE,Bangalore
Prof. Sushanta Kumar Panigrahi
IIMT, Bhubaneswar
Prof. Mritunjay Sharma
IIMT, Bhubaneswar

Organizing Committee:

Bibhu Prasad Mohanty
Pritika Mohanty
Bijan kumar Barik
Ganeswar Mishra

First Impression : 2012

International Joint Conference

On

EMERGING INTELLIGENT SUSTAINABLE TECHNOLOGIES

(EISTCON-2012), Volume-III

DISCLAIMER

The authors are solely responsible for the contents of the papers complied in this volume. The publishers or editors do not take any responsibility for the same in any manner. Errors, if any, are purely unintentional and readers are requested to communicate such errors to the editors or publishers to avoid discrepancies in future.

ISBN : 978-93-81693-76-6

Published by :

IPM PVT. LTD., Interscience Campus
At/PO.: Kantabada, Via: Janla, Dist: Khurda, Pin- 752054
Publisher's Website : www.interscience.in
E-mail: ipm.bbsr@gmail.com

Typeset & Printed by :
IPM PVT. LTD.

TABLE OF CONTENTS

Editorial

In the pursuit of scientific excellence human society is accelerating the pace of technological developments in every discipline. In the 21st century the most critical concern of every research is to address the environmental and sustainability issues. Specifically in the areas like Power and Electronics engineering where energy is an integral component of analysis, scholars and practitioners must exhibit a thrust on energy aspect of their research interest. The major drivers of a new paradigm can be attributed to:

- Environmental pressures
- Customer pressures
- Financial pressures
- Technology developments

There are several reason why the traditional technologies are not appropriate anymore for current usage:

- Do not meet current environmental pressures
- Cannot readily meet customer pressures
- Are no longer the most appropriate solutions for developing countries
- Have long term paybacks which do not meet current financial restraints
- Do not allow for innovative solutions

The new electricity production technologies consist of:

- Development of gas infrastructures
- Application of gas fired generation closer to loads
- Combined heat and power Embedded generation
- Development of micro turbines
- Advancement of fuel cell technologies
- Development of renewable sources
- Development of energy storage technologies
- Development of photovoltaic technologies
- Nuclear generation

For a technological point of view it means that a user centric approach is needed (smart grid). Furthermore to enable secure and sustainable solutions integration of utility infrastructures and super grids are needed. For the next decades a detailed asset management needs to be done in order to determine the efficient investments, sustainable utilization and precise removal to have an effective hybrid network consisting of old and new technologies. Another aspect that needs to be discussed is energy conservation which refers to the preservation of fossil energy sources (oil and gas). For that restrictions need to be imposed. Also it needs to be pointed out that we are trying to conserve raw materials and the environment. Saving of energy will only assist but will not be part of energy conservation.
From a material point of view it will require a greater cooperation between technology equipment and material developers, identification of idealized material requirement for new technologies and a decrease of development periods. The following aspects for material developments need to be considered:

- High temperature superconducting compounds
- Polymers
- Ceramics
- Semiconductors

- Liquid insulation
- Gas insulation etc

The word Computing has obtained a different connotation in the context of advance studies in Computer Science. As the computational complexities are increasingly influence the researchers in several multidisciplinary studies the approach should to develop advanced techniques in easing an user-friendly end product. In the latter half of the twentieth century witnessed incremental innovations in the field of computational science. The growth of such a distinctive field dates back to 1960 when John McCarthy opined "computation may someday be organized as a public utility". The present scenario of computer vision and cloud characteristics were explicitly described by the scholarly bibliography by Canadian technologist Douglass Parkhill's "The Challenges of the Computer Utility". Other scholars have shown that cloud computing's roots go all the way back to the 1950s when scientist Herb Grosch (the author of Grosch's law) postulated that the entire world would operate on dumb terminals powered by about 15 large data centers.

The mission of Roadmap of future is to provide a practical, independent and objective analysis of pathways to achieve low-carbon economies in the world in line with the energy security, environmental and economic goals.

The conference is designed to stimulate the young minds including Research Scholars, Academicians, and Practitioners to contribute their ideas, thoughts and nobility in these emerging disciplines. Even a fraction of active participation deeply influences the magnanimity of this international event. I must acknowledge your response to this conference. I ought to convey that this conference is only a little step towards knowledge, network and relationship

I sincerely thank to the board of editors for their generous gifts of time, energy and intellect. I salute to the technical committee, organizing committee and publishing house for their unbroken professional commitments in bringing this proceeding.

I hope this gracious piece of effort will earn a significant appreciation from the readers to whom I owe indebtedness. I must welcome constructive feedbacks for future development.

With Warm Regards

Editors
Prof. (Dr.) Srikanta Patnaik
President, IRNet India and Chairman IIMT
Intersceince Campus, Bhubaneswar
Email: srikantapatnaik@hotmail.com

&

Dr. K. Shanmukha Sundar
Prof & HOD.
Electrical & Electronics Engineering Department,
Dayananda Sagar College of Engineering
Bangalore-560078, India
E-mail: bonniekhs@gmail.com

Cyber Security Solutions for DLMS Meters Using GSM/GPRS Technology

Anju P Rajan Mathew[1], A. AjilAylwin[2] & Shaileshwari M U[3]

[1&2] Department of CSE, The Oxford College of Engineering, Bangalore
[3]Engineering Officer Grade 2, Central Power Research Institute, Bangalore, India,
E-mail : anjuplamthottam@gmail.com, aajilnila@gmail.com & shaileshwari@cpri.in

Abstract **-** The Smart meters are used in the areas of generation, transmission, distribution and consumption. The capabilities of smart meter systems and grid networks, such as distributed intelligence and broadband capabilities can greatly enhance efficiency and reliability, but they may also create much new vulnerability if not deployed with the appropriate security controls. Much of the technology currently in use by the meters are outdated and in many cases unreliable. A system architecture implementing should recognize security threats and capture events that result not from external threats but from internal mistakes, with human error being a more common occurrence. An effective security approach enhances reliability because some security failures might be people failures, while others might be equipment failures, might be due to natural causes or might be deliberate. A simple perimeter defense is not sufficient; monitoring, both for events and physical actions, is required to bring the benefits of smart meters with minimal risk to this vital part of the infrastructure of modern life.

Keywords-*DLMS/COSEM, SmartMeters, Application Association (AA),GPRS, GEA, Signaling System 7 (SS7), TMSI, IMSI.*

I. INTRODUCTION

DLMS/COSEM specification specifies a data model and communication protocols for data exchange with metering equipment. It follows a three-step approach:

- Step 1. Modelling: This covers the data model of metering equipment as well as rules for data identification. The data model provides a view of the functionality of the meter, as it is available at its interface(s). It uses generic building blocks to model this functionality. The model does not cover internal, implementation-specific issues.
- Step 2, Messaging: This covers the communication services and protocols for mapping the elements of the data model to application protocol data units (APDU).
- Step 3, Transporting: This covers the services and protocols for the transportation of the messages through the communication channel.

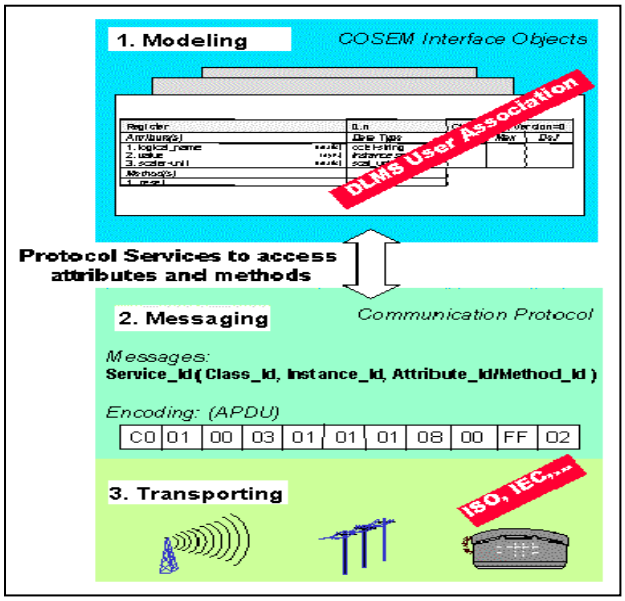


Fig.1 The three steps approach of DLMS Modelling – Messaging – Transporting

A. *DLMS Based Smart Meter Overview*

The liberalized energy market requirements given by DLMS meters are;

- Unbundling of monopolistic utilities.
- Introduction of competition in all activities: – generation – transport – supply – customer management -meter operation –meter reading – meter data management.
- Geographical dispersion, volatile customer base.
- Multi-energy - multi-user - multi-vendor environment.
- Need selective and secure access to data and interoperability.

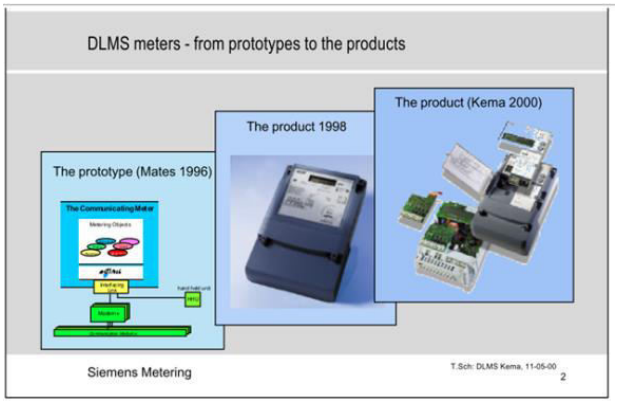


Fig.2 DLMS meter

B. *Why DLMS*

DLMS is comparable to a set of rules or a common language, on which the various operators have agreed. The DLMS-protocol enables the integration of energy meters with data management systems from other manufacturers. This secures that the energy supplier gets the full advantage of the meter functions. The utility that has invested in a smart metering solution pulls an enormous amount of information out of the meters ie;information that can be used for a lot more than billing purposes such as:

- Load control
- Development of tariff models for special customer segments

DLMS-protocol enables Energy trade and has Low communication overhead. When using it only necessary data reaches utility software.

II. DATA ACCESS SECURITY IN DLMS METER

Data access security concerns role based access to data in a DLMS/COSEM device. It is managed by the Association LN / Association SN objects. Each COSEM server i.e., a logical device may support AAs with various clients, each having a different role, and with this, different access rights. Each AA is identified with a pair of lower layer addresses. Each Association object provides a list of objects visible in that particular AA and the access rights to their attributes and methods. To be able to access data, the client must be properly authenticated. Upon AA establishment, an authentication context is negotiated between the client and the server. This specifies the required authentication of the peers, and, where needed, the security algorithm to verify the authentication.

A. *Authentication In Data Exchange*

The security of exchanges includes authentication of the Client and the Server, confidentiality of exchanged datacontrol of access to the variable objects of the Server. Authentication enables the Server to control the identity of the Client in order to provide with proper access rights. When this authentication is mutual, the Client can also control the Server. The data confidentiality is taken care for authorized readings only and data exchange is protected. Masking options of the message and time provides additional and adequate security.

The standard supports two security schemes that are defined in ICS:

- Basic security
- Advanced security.

B. *Basic security (Authentication)*

Basic security is an access control which provides authentication as addressed in COSEM specification. In order to provide different levels of security for authentication support, COSEM specifies three levels of authentication securities:

- No authentication (lowest level) security.
- Low level, password based authentication security (LLS) identifying only the client.
- High-level, four-pass authentication security (HLS) identifying both the client and the server.

The meter supports three associations in the Management Logical Device:

- Public Client association (PC)
- Meter Reader association (MR)
- Utility Settings association (US)

C. No authentication security

This level provides access to the server without any authentication during sign on the access to the server is read only for PC association.

D. Low-Level Security

Low level authentication security offers adequate security to avoid eavesdropping and message (password) replay during data transfer. This level of security is mandatory in MR association which will have a password for data download. The association objects provide an attribute called 'LLS_secret' to change the secret (low level security password) only by the authorized client.

Verification of Low-Level Security: Client transmits a 'secret' to the server, by using the 'Calling Authentication Value' parameter of the COSEM-OPEN. Request service primitive of the client application layer.The server verifies the received 'secret' with the corresponding client identification and the association is established, if the received 'secret' is valid.

E. High Level Security

High-level authentication security is typically used when the communication channel offers extrinsic security and suitable measures to be taken to avoid eavesdroppers and message (password) replay. This level of security is mandatory for the US mode of association. HLS mechanism defines a 4-pass sign-on scheme where the client and server exchange challenges and then reply to the challenges with a processed response. The processing performed on the challenges is an encryption using a secret key.

Pass1: The client transmits 'challenge' to the server (CtoS).

Pass2: The server transmits 'challenge (as acknowledgement)' to the client (StoC).

Once the Pass 2 is also valid, the association is formally established between the client and the server but the access of the client is restricted to the method "reply to HLS authentication" of the current "association" object.

Pass3: The client processes StoCin a secret way. The result of (StoC) is sent back to the server. The server checks the result of correct processing and if correct the server accepts the authentication of the client.

Pass4: If the client is authenticated, the server processes (CtoS) in a secret way. The result of (CtoS) is sent back to the client. The client checks the result of the correct processing and if correct, the client accepts the authentication of the server.

Pass3 and Pass4 are supported by the method"reply_toHLS_authentication"of the association object(s). If both passes are successfully executed, then full access is granted according to the current association. Otherwise, either the client or the server aborts the association. Inaddition; the association object provides the method to change the HLS 'secret' (e.g. the encryption key): change_HLS_secret.

F. Advanced Security (Data Security)

Transport of data is done in secret way e.g. the encryption key. Encryption is the conversion of data into a form, called a cipher text. Chipper text cannot be easily understood by unauthorized client. It is mostly used for data security purpose. The proposed standard for encryption is AES GCM for ICS.

III. GRID NETWORK VULNERABILITIES

The listed vulnerabilities are not necessarily ordered according to severity, which is affected by the particular utility type, infrastructure, potential attacker profile and many other factors that need to be determined in the general risk assessment process.

A. Network Management from Remote Nodes

Each meter is a node in the Smart Grid network. Although the processesbeing executed on the network require only data to be read and commands tobe sent to the meter, the management applications and services remainexposed and available for all the nodes.The practical implications of this scenario is that without explicit constraints,an attacker who uses the communication module of the smart meter cancause network-wide changes, ranging from disrupting the communication flowto rerouting all the traffic to his node for later manipulation.

B. Lack of Authentication

Security [6] has encountered numerous meters that didn't have anyauthentication or encryption support. This design flaw makes it possible for anattacker to impersonate the control center and send unauthorized commandsto meters or read metering data. The consequence of a successful attack onmeters with disconnection capabilities is particularly destructive. It should be noted that although some of the metering protocols supportencryption, which can be viewed as a network access password, most of thedeployments we've encountered so far did not enable these features. Sinceevery metering standard includes support for "no encryption" or "noauthentication", it usually poses too great a temptation for the integrationteams which prefer to choose these settings in order to avoid additionaldeployment problems.

C. *Authentication Bypass*

Several metering protocols (DLMS, IEC 60870-5-102) implementation havefunctions to read metering data which do not require a password, andconfiguration/disconnect functions that require the operator password. Twometers that we audited retrieved the password for the restricted functionsusing the unprotected read function. This implementation makes theauthentication/password protection completely useless.

D. *Slave Meter Data Tampering*

The protocol used for communication between the master (smart) meter andthe slave meter is usually considered of lesser importance as its impact isrestricted to the single customer household. Although this is generally correct,from a risk management point of view it is important to identify and address asituation where a cheap "man in the middle" device is inserted between themaster and slave meters which lowers the usage reading by a constantdivision.

This manipulation is both very hard for the utility to identify and canhappen in a large scale if a criminal party decides to mass produce andmarket these devices – much like pirate cable set-top boxes / satellitedecoders.

E. *Slave Meter Unauthorized Disconnection*

Some slave meters support disconnection of the customer upon receiving arequest from the master meter. Normally the associated risk is minimal as ifan attacker was to disconnect the slave meter, as these meters are commonlyconnected by wire to the master meter the physical presence is required andtherefore disconnection could be achieved by bringing a hammer. Thisassumption causes to set low security settings to this communication channel,as it is perceived as non-critical. Unfortunately, some of the meteringprotocols used between meters are wireless (e.g. WMBUS, Z-Wave) making itpossible for an attacker with a potent transmitter to send a disconnect signalto multiple customers, especially in crowded urban areas. The attacker will notneed receive the data back from the meters to issue this command.

F. *Insecure Protocol Implementation*

Meters from a variety of vendors that were audited by Security [6] were foundto improperly handle malformed requests. When a meter firmware makescertain assumptions regarding the data it receives, and in particular themaximum size of each message type, it may be vulnerable to a very well-known attack condition named Buffer Overrun/Overflow Vulnerability. Thisvulnerability may allow the attacker to cause system instability or freeze,change values of parameters which are saved in the memory stack or evenexecute arbitrary code. In most of the meters and RTUs that were audited byour "red team", such a condition was identified and exploited.

G. *Firmware Upgrade Vulnerabilities*

Firmware upgrades are a double edged sword. The existence of the capabilityto remotely upgrade the meter firmware is of crucial importance – as securityexperts like to repeat a well-known, and true, mantra that "what is consideredsecure today may be proven otherwise tomorrow". There's no assurance thata new unforeseen attack will successfully compromise a meter model and soin order to be able to respond the operator must have the ability to securelyupdate the meter firmware to upgrade as many meters as it can before theyare compromised.The other side of firmware upgrades is that they serve as a powerful tool forattackers, if they can be abused. For example, an attacker who can push hisown firmware to other meters can execute a disconnect action and then makethe meter completely unresponsive till it is returned to the manufacturer, thusmaking it impossible for the network operator to reverse his actions.To conclude, it is crucial to have a remote firmware upgrade capability, butone that was designed with security in mind and audited thoroughly byexperts.

H. *Input Validation*

The all-too-familiar security problem of input validation, which is unfortunatelyquite common in control systems, was found to exist in Smart Grid meters andservers as well. Should an attacker be able to broadcast malformedmessages to a node on the Smart Grid (which we elaborated on why that cannormally easily be done) it will have a relatively high success probability tocause the node to fail. The failure is a result of assuming that the datareceived is in the expected message format, whereas when a malformedpacket is parsed it causes an exception that may even lead to arbitrary codeexecution.

IV. GPRS SECURITY ARCHITECTURE

In order to meet security objectives, GPRS employs a setof security mechanisms that constitutes the GPRS security architecture. Most of these mechanisms have beenoriginally designed for GSM, but they have been modified to adapt to the packet-oriented traffic nature and theGPRS network components. The GPRS security architecture, mainly, aims at two goals: a) to protect the networkagainst unauthorized access, and b) to protect the privacyof users. It includes the following components:

- Subscriber Identity Module (SIM)
- Subscriber identity confidentiality
- Subscriber identity authentication
- User data and signaling confidentiality between the MS and the SGSN
- GPRS backbone security

A. Subscriber Identity Module – SIM

The subscription of a mobile user to a network is personalized through the use of a smart card named SubscriberIdentity Module (SIM). Each SIM-card is unique andrelated to a user. It has a microcomputer with a processor, ROM, persistent EPROM memory, volatile RAM and an I/O interface. Its software consists of an operatingsystem, file system, and application programs (e.g., SIMApplication Toolkit). The SIM card is responsible for the authentication of the user by prompting for a code (Personal Identity Number PIN).

A serious weakness of the GPRS security architecture isrelated to the compromise of the confidentiality of subscriber identity. Specifically, whenever the serving network (VLR or SGSN) cannot associate the TMSI withthe IMSI, because of TMSI corruption or database failure, the SGSN should request the MS to identify itself bymeans of IMSI on the radio path. Furthermore, when theuser roams and the new serving network cannot contactthe previous (the old serving network) or cannot retrievethe user identity, then, the new serving network shouldalso request the MS to identify itself by means of IMSIon the radio path. This fact may lead an active attackerto pretend to be a new serving network, to which theuser has to reveal his permanent identity. In addition, inboth cases the IMSI that represents the permanent useridentity isconveyed in clear-text over the radio interfaceviolating user identity confidentiality.

B. Subscriber Identity Authentication

A mobile user that attempts to access the network mustfirst prove his identity to it. User authentication protects against fraudulent use and ensures correct billing.GPRS uses the authentication procedure already definedin GSM with the same algorithms for authentication andgeneration of encryption key, and the same secret key, Ki. However, from the network side, the wholeprocedure is executed by the SGSN (instead of the basestation) and employs a different random number (GPRS-RAND), and, thus, it produces a different signed response(GPRS-SRES) and encryption key (GPRS-Kc) than theGSM voice counterpart.

The authentication mechanism used in GPRS also exhibits some weak points regarding security. More specifically, the authentication procedure is one-way, and, thus, it does not assure that a mobile user is connected to anauthentic serving network. This fact enables active attacks using a false base station identity. An adversary, who has the required equipment, may masquerade as alegitimate network element mediating in the communication between the MS and the authentic base station. Thisis also facilitated by the absence of a data integrity mechanism on the radio access network of GPRS, which defeats certain network impersonation attacks. The resultsof this mediation may be the alternation or the interception of signaling information and communication dataexchanged.

C. Data and Signaling Protection

User data and signaling protection over the GPRS radioaccess network is based on the GPRS ciphering algorithm (GPRS-A5), which is also referred to as GPRS Encryption Algorithm (GEA) and is similar to the GSMA5. Currently, there are three versions of this algorithm: GEA1, GEA2 and GEA3 (that is actually A5/3), whichare not publicly known, and, thus, it is difficult to performattacks on them. The MS device (not the SIM-card) performs GEA using the encryption key (GPRS-Kc), since itis a strong algorithm that requires relatively high processing capabilities. From the network side, the serving SGSNperforms the ciphering/deciphering functionality protecting signaling and user data over the Um, Abis, and Gbinterfaces.

An important weakness of the GPRS security architectureis related to the fact that the encryption of signaling anduser data over the highly exposed radio interface is notmandatory. Some GPRS operators, in certain countries,are never switch on encryption in their networks, sincethe legal frameworks in these countries do not permit that. Hence, in these cases signaling and data traffic are conveyed in clear-text over the radio path. This situation isbecoming even more risky from the fact that the involvedend-users (humans) are not informed whether their sessions are encrypted or not.

D. GPRS Backbone Security

The GPRS backbone network includes the fixed networkelements and their physical connections that convey userdata and signaling information. Signaling exchange inGPRS is mainly based on the Signaling System 7 (SS7)technology, which does not support any security measure for the GPRS deployment. Similarly, the GTP protocol that is employed for communication between GSNsdoes not support security. Thus, user data and signalinginformation in the GPRS backbone network are conveyedin clear-text exposing them to various security threats.In addition, inter-network communications (between different operators) are based on the public Internet, whichenables IP spoofing to any

malicious third party who getsaccess to it. In the sequel, the security measures appliedto the GPRS backbone network are presented.

Based on the analysis of the GPRS security architectureit can be perceived that the GPRS securitydoes not aim at the GPRS backbone and the wire-lineconnections, but merely at the radio access network andthe wireless path. Thus, user data and signaling information, conveyed over the GPRS backbone, may experience security threats, which degrade the level of securitysupported by GPRS. In the following, the security weaknesses of the GPRS security architecture that are relatedto the GPRS backbone network for both signaling anddata plane are presented and analyzed.

V. CONCLUSION

System security and Data security is a critical issue today. A comprehensive architecture with security built in from the beginning is necessary. Grid security involves an architecture that includes security from the beginning, consists of more than just protective devices such as firewall, and engages processes as well as products. A simple perimeter defence is not sufficient; monitoring, both for events and physical actions, is required to bring the benefits of grid with minimal risk to this vital part of the infrastructure of modern life. GPRS promises to benefit network users greatly by providing always on higher bandwidth connections than are widely available today. In order to be successful, data connections must be secure and be available all the time from anywhere. The maturity of security in the air interface and the low bandwidth available limits the effectiveness of the Network Station as the source of attacks. With the increase in the use of wireless media, security problems of confidentiality, integrity, and authentication are also increasing. The weak points of the GPRS security architecturemay lead to compromises of end-users and network security of the GPRS system. The proposed enhancements can be easily integrated in the existing GPRS infrastructure, minimizing the required changes.

ACKNOWLEDGMENT

Authors gratefully acknowledge the officers of Utility Automation Research Center in Central Power Research Institute, Bangalore for the technical support provided by them. We would like to thank all the lectures in the Dept.of CSE from The Oxford College of Engineering whose support made a difference in research. We also thank the HOD of CSE Dr. R. J. Anandhi for her encouragement and support in our endeavors.

REFERENCES

[1] BLUE, GREEN & YELLOW BOOK FROM DLMS UA.

[2] SOMOGYI Tibor, "DLMS-the Application Protocol for communicating meters", *Essen*, April, 1997.

[3] *Report of High Level Committee on Standardization of Meter Protocol*,Central Electricity Authority New Delhi, December 2008.

[4] Rupok, M.S.A,"Design and implementation of a novel remote metering system using USB GPRS/EDGE modem",Electronics Computer Technology (ICECT), 2011 3rd International Conference on,July 2011

[5] GPRS Security Threats and Solutions,March 2002[Online], Available:http:// www. netscreen.com /us/ en/training/ GPRS Security.pdf.

[6] The Dark Side of the Smart Grid -Smart Meters (in) Security, 2011[Online], Available:http:// www.c4security.com/c4/TheDarkSideofheSmart GridSmartMeters (in) Security.pdf.

[7] Wilayat Khan and HabibUllah," Authentication and Secure Communication in GSM, GPRS, and UMTS Using Asymmetric Cryptography",*IJCSI International Journal of Computer Science Issues*, Vol. 7, Issue 3, No 9, May 2010.

[8] Report to NIST on Smart Grid Interoperability Standards Roadmap EPRI, Jun.17,2009 [Online]. Available:http://www.nist.gov/smartgrid/Interim SmartGridRoadmapNISTRestructure.pdf.

❑❑❑

Implementation of Red Algorithm On FPGA For High Speed Pupil Isolation

Banushri .S, M.Z.Kurian & M. Nagaraja

Department of Electronics and Communication
Sri Siddhartha Institute of Technology, Tumkur, Karnataka, India
E-mail : banushri14@gmail.com

Abstract **-** A biometric system provides automatic identification of an individual based on a unique feature or characteristic possessed by the individual. Iris recognition is regarded as the most reliable and accurate biometric identification system available. In the Department of Defense and in the Department of Homeland Security, iris recognition plays a very important role. At first image acquisition is done, the image is subjected to the calculation of histogram and then binarization is conducted. The performance of iris recognition system highly depends on edge detection. The Canny Edge Detector is one of the most commonly used image processing tools; it detects the edges in a very robust manner. An algorithm that is both accurate and fast in a hardware design that is small and transportable is crucial to the implementation. As a part of an ongoing effort to meet these criteria, the Ridge-Energy-Direction(RED) algorithm were proposed which reduces the effects of illumination, since only direction is used. We show a significant speed-up of pupil isolation by implementing this portion of the algorithm on a Field Programmable Gate Array (FPGA).

Keywords: *Iris recognition, RED algorithm, FPGA, Canny edge detection.*

I. INTRODUCTION

The term biometrics is derived from the Greek words bio (life) and metric (to measure). It refers to the automated technique of measuring a physical or behavioral characteristic of an individual by using the technologies that automatically authenticate, identify or verify. In essence, biometrics allows confirmation or identification of WHO YOU ARE. To support WHAT YOU KNOW (a pin or password) and WHAT YOU HAVE GOT.

Nowadays, security is one of the important factors in the field of information, business, e-commerce, military and etc. For this reason, personal identification has become a significant topic. Some previous methods of identification such as PIN (personal Identification Number), password, ID card and signatures that are used widely, have some drawbacks [1 and 2]. ID card or PIN can be stolen, password may be forgotten and signatures can be imitated. The purpose of 'Iris Recognition', a biometrical based technology for personal identification and verification, is to recognize a person from his/her iris prints. In fact, iris patterns are characterized by high level of stability and distinctiveness. Each individual has a unique iris, the difference even exists between identical twins and between the left and right eye of the same person.

John Daugman created a strong algorithm which was based on Gabor wavelets for iris detection. An alternate design, called Ridge Energy Direction (RED) was proposed which was based on spatial domain directional filters. This application has a great military interest, and implementing this on a small system on a chip design using a Field Programmable Gate Array (FPGA) would be ideal for carrying into the field or storing in a backpack. FPGA devices also provide an attractive solution to computationally intensive applications because of their high density, high performance and complete configurability to support specific applications.

An FPGA chip offers a combination of the flexibility of general purpose computers and hardware-based real time processing of application specific integrated circuits (ASICs) .An architecture design for FPGA technology can fully exploit the data and I/O parallelism in most of the image processing applications. The main objective of this project is to build each section of the RED algorithm using the most efficient method and combine them to create a single fast and accurate system for iris recognition.

To achieve this end state there are several steps. Firstly, from the infrared image of the eye of a compliant participant, the iris must be extracted as

accurately as possible. There are several factors that can cause interference such as the presence of the eyelid and eyelashes, reflections, and different angles of the eyeball. The application is to isolate and extract the pupil from within the iris. One of the means to do this is the RED algorithm. Using an FPGA this paper speeds up the process of finding the inner border of the iris.The comparison programs are used to show the speed-up which is achieved on the FPGA over the CPU and GPGPU. Initially, the CHT has been used for the FPGA application; however we found that the transform is not ideal when implemented in hardware since a counter has to be applied to a large array which requires hundreds of reads and writes per clock cycle. This implementation cannot be speed up easily and it also requires a large amount of memory. Instead of that, the method used was establishing a search space where the pupil is expected to be found and building a circle around each potential center-point in the search space, repeating with different radii.

II. EXISTING SYSTEM

Biometrics is playing very important role in day to day life. It is used for various security solutions. Biometric technologies are becoming the foundation of an extensive array of highly secure personal identification and verification solutions.Iris has some advantages over other biometrics [6]. The iris is a protected, internal organ, and externally visible. The texture pattern of an iris is highly stable with age and health condition. Iris data is non-identical for left, right eyes and for twins also [7].

No need for a person to identified to touch any equipment. It is evident that iris and fingerprint recognition have high distinctiveness, permanence, and performance than any other biometric systems. Iris recognition is one of the very important biometric applications.

Classical approaches for the iris locating were Daugman approach and Hough transforms based approach. It was carried out using software which takes long time for its execution. To increase the performance and time for the same recognition system we need to improve or introduce hardware architecture for the same. Hence we are developing high speed pupil isolation which will definitely improve the system. In the recent years many algorithms are proposed.

III. PROPOSED SYSTEM DESIGN

The block diagram of the proposed system is shown in Fig .1.Firstly the image is been acquired, experiments are performed using iris images obtained from CASIA database (Institute of Automation, Chinese Academy of Science s) the image will be usually in RGB format, the RGB format image will be converted into binary image then after edge detection will be done. The edge detection is carried out by canny edge detection. After that implementation of the RED algorithm is done for isolation purpose. We are going to implement RED algorithm on an FPGA for achieving high speed. The Ridge Energy Direction algorithm is used for Iris recognition which requires several steps including edge detection, iris isolation, unwrapping, and search. And after implementing the RED algorithm, result is the isolated pupil. Simulation is done in MATLAB. A MATLAB is a high-level technical computing language and interactive environment for algorithm development, data visualization, data analysis and numeric computation. And application of MATLAB is an easy and efficient tool in image manipulation. And finally the selected objective is to perform verification of the developed module in terms of high speed and also to obtain better accuracy.

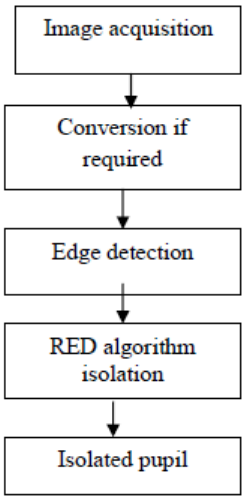


Fig.1. Main block

IV. SYSTEM IMPLEMENTATION

Several steps are needed to be carried out for converting the original iris image to binary image which are discussed below

A. Calculation of histogram

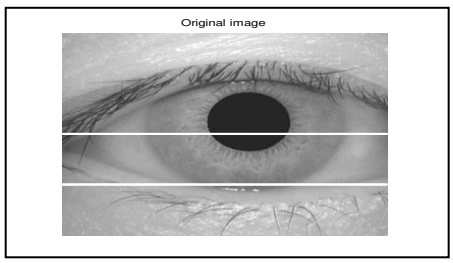


Fig.2.Original image

For the purpose of converting the original image (Fig.2.) to the binary image, we should choose a reasonable threshold value. At first, we analyze the histogram (Fig .3.) of the original iris image.

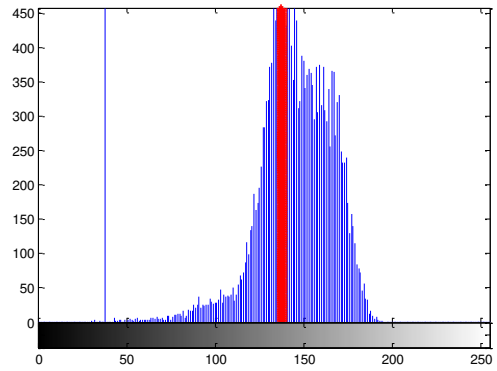


Fig.3.Histogram

Histogram is a graphical representation of data distribution. Histogram is used to enhance the image contrast or to determine the threshold values. Histogram is a useful tool to analyze the brightness and contrast of an image. It shows how the intensity values of an image are distributed and the range of brightness from dark to bright. An image can be enhanced by remapping the intensity values using the histogram.

B. Binarization

A binarization is a process of converting a color image is converted into a binary image. A binary image is a digital image that has only two possible values for each pixel. Typically the two colors used for a binary image are black and white .Binary images are also called bi-level or two –level. The image pixels have two intensity values- 0 for black and either 1 or 255 for white.

B. Algorithm for binarization:

Step1: Consider the original image.

Step2: Each pixel is constituted of mainly 3 colors namely red, green and blue and now in this step we are going to check values of each color.

Step3: If the value of each color in the each pixel<127, the pixel value will be made 1.

Step4: Otherwise set to 0.

Step5: Binarized image is obtained.

C. Applying of Morphological operations

Morphological operations are used to understand the structure or form of an image. This usually means identifying objects or boundaries within an image. Morphological operations are usually performed on binary images where the pixel values are either 0 or 1. Morphological operation is of mainly two types called dilation and erosion. Dilation is an operation that allows a black pixel to remain black only if all of its neighbors are black. Dilation is used for expanding shapes and filling holes, gaps. Erosion is an operation that allows white pixel to remain white in the output image is if all of its neighbors are white. The effect on a binary image is to diminish, or erode, the edges of a white area of pixels. Erosion is used for shrinking shapes and removing bridges, branches and small protrusion. After the application of erosion and dilation for an iris image, we have defined a bounding box (BB) which covers the iris region of the image. The bounding box is merely the coordinates of the rectangular border that fully encloses a digital image when it is placed over a page, a canvas, a screen or other similar bi-dimensional background. Image within the BB is used to find the edge points and then we filter out some of the edge points and edge detection is employed for the boundary detection.

D. Applying of canny edge detector

Edge detection refers to the process of identifying and locating sharp discontinuities in an image. The discontinuities are abrupt changes in pixel intensity which characterize boundaries of objects in a scene. Several algorithms exist, in that one of the standard edge detection methods is the one developed by John F. Canny (JFC).

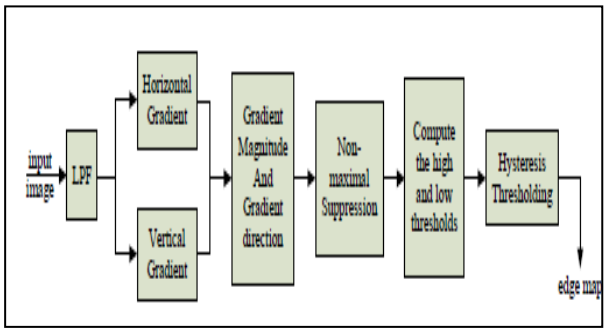


Fig.4.Canny Edge Detection Algorithm

Fig.4. shows the block diagram of canny edge detector.

Canny edge algorithm steps:

Step 1: In order to implement the canny edge detector algorithm, a series of steps must be followed. The first step is to filter out any noise in the original image before trying to locate and detect any edges. And for this, the Gaussian filter can be computed using a simple mask.

Step 2: After smoothing the image and eliminating the noise, the next step is to find the edge strength by taking the gradient of the image. The Sobel operator uses a pair of 3x3 convolution masks, one estimating the gradient in

the x-direction (columns) and the other estimating the gradient in the y-direction (rows). The magnitude, or edge strength, of the gradient is then approximated using the formula:

$|G| = |Gx| + |Gy|$

Step3: The direction of the edge is computed using the gradient in the x and y directions. The formula for finding the edge direction is just:

Theta = invtan (Gy / Gx)

Step 4: After the edge directions are known, non maximum suppression now has to be applied. Non maximum suppression is used to trace along the edge in the edge direction and suppress any pixel value that is not considered to be an edge. This will give a thin line in the output image.

Step 5: Finally, hysteresis is used as a means of eliminating streaking. Streaking is the breaking up of an edge contour caused by the operator output fluctuating above and below the threshold. The application of the canny edge detection algorithm is shown in Fig.5.After that we need to remove the noise which is shown in Fig.6.

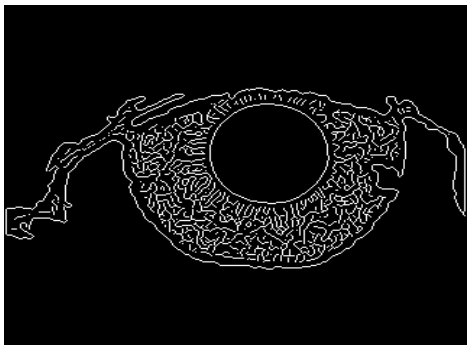


Fig.5. canny edge detection

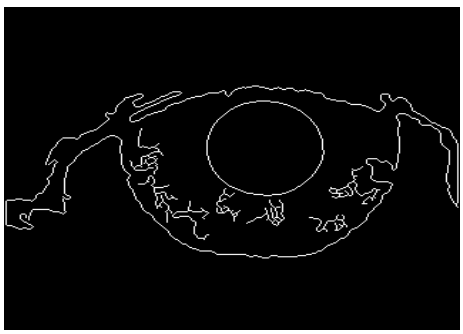


Fig.6. Removal of noise.

E .Implementation Aspects

VHDL is used as the hardware description language because of the flexibility to exchange among environments. The code is pure VHDL that could easily be implemented on other devices, without changing the design. The software used for this work is Synthesis Tool Xilinx 12.2. This is used for writing, debugging and optimizing efforts, and also for fitting, simulating and checking the performance results using the simulation tools available on Modelsim6.3c and MATLAB. Modelsim is verification and simulation tool for VHDL, verilog, system verilog and mixed language design.

V. RED ALGORITHM

A new iris recognition algorithm was presented that served an alternative to many of the commercial algorithm such as Daugman algorithm [5]. It incorporated local statistical analysis in segmentation and used the direction of the ridge pattern that appear in the unwrapped iris in the feature extraction .The Ridge Energy Direction algorithm was created at the United States Naval Academy .The RED algorithm has been shown to have a comparable performance to state-of-the art algorithm. Iris recognition requires several steps such as edge detection, iris isolation, unwrapping and search. The RED algorithm is based on the prominent direction of the ridges. For this project, we assume the full iris is shown orthogonally, so that the iris and pupil can be approximated by a circle. Previous and later steps can neutralize many deviations from this assumption, such as eyelid and eyelash interference or slightly non-orthogonal images. The RED algorithm was initially implemented in C++. In [5], the author has indicated hardware acceleration as future work. In [6], the authors have implemented the template generation segment of the RED algorithm on an FPGA achieving a 324 times speed-up. Our paper realizes the pupil isolation segment of their algorithm achieving a 32 times speed-up.

A. Creating a Circle with a Given Center-point and Radius.

Given an (x, y) coordinate from the search area and an even radius between 10 and 126 pixels, a list of address points for a circle with center-point (x, y) and radius r is created. The circle plot area is a 256 by 256 pixel space which is taken from the edge detected image. The address points are computed by implementing a version of the CORDIC algorithm. The CORDIC module (Fig. 7) is copied 91 times in parallel to create the first 91 points on a circle from t=0 to t=90 in one clock cycle.

$$x_0 = x_{center} + r\cos(t)$$
$$y_0 = y_{center} + r\sin(t)$$

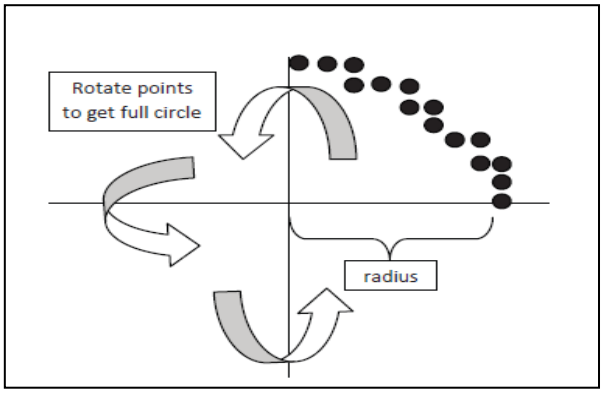


Fig.7.CORDIC Module in one clock cycle

For implementing on the hardware we need the following requirements such as DSP supported FPGAs which is readily available and it contains a substantial number of technology-optimized and built-in components that are frequently used in DSP applications such as multipliers, Multiply-and-Accumulator (MAC) units, and embedded memory. In the memory controller these points will be used to form the full circle. After the 91 points are created they go into an array and are compared to remove any repeat points created due to rounding. Once this is complete, the points are ready to move into the memory controller.

B. The memory controller.

Memory controller initiates the computational process by issuing a control signals and start reading the pixel location from the list. Copies of binary images are stored in the embedded memory structures for fast and parallel accessing. The target FPGA (Stratix IV family) consists of columns of M9K and M144K memory blocks that can be configured as dual-port Random Access Memory (RAM) modules, which support two read, two write or one read one write operations simultaneously [7]. These embedded memory blocks have been optimized to support real-time processing applications operation. Each M4K block is configured as an 8192×1 RAM module. Eight embedded M4K blocks are used to form a complete 256×256 binary image. Each pixel in the edge detected image is represented by one bit that can be consecutively laid out in the addressable space. This consecutive mode is the most efficient way to utilize all the memory bits. Each bit can be addressed by calculating the offset based upon the pixel location. Fig .8.shows the block diagram for a memory controller.

The computing time is to be found and compared with the CPU and GPGPU (General Purpose Graphics Processing Unit). The FPGA implementation shows a great speed-up and it is a better platform for RED algorithm.

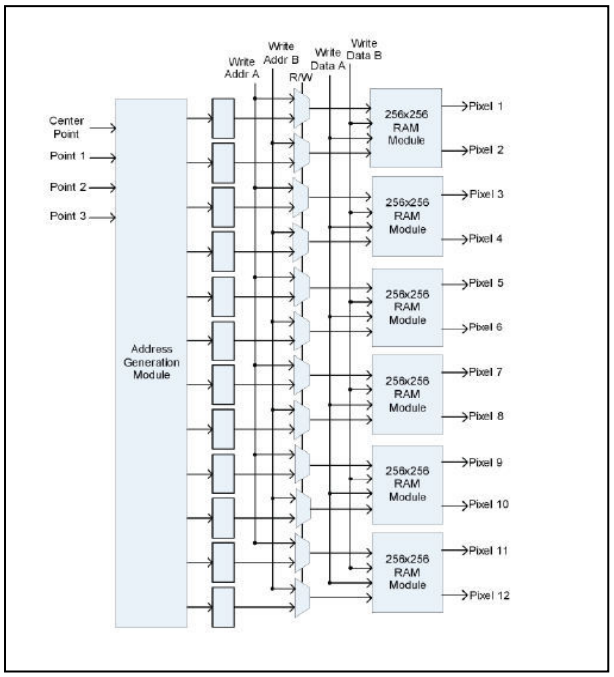


Fig.8. The memory controller

VI. CONCLUSION

Iris recognition has a long and colorful history. A new iris recognition algorithm was presented that can serve as an alternative to other commercial algorithms. This paper implements a portion of red algorithm on FPGA achieving a greater accuracy and efficiency. This method carries with it several assumptions such as it is assumed that the iris images are orthogonal, such that the eye is looking directly. In conjunction with this, it is assumed that the pupil and the limbic boundary of the iris is circular, which is not always accurate. Overall, the algorithm has several areas that can be addressed to improve performance; these are discussed in the next section.

VII. FUTURE WORK

The hardware implementation of RED algorithm is still under development. Several modifications are scheduled for near-term improvement, some of which have been alluded to in this paper .The next implementation will include the canny edge detection algorithm on FPGA as well. There are other portions of RED that could be improved using hardware design. Specifically, in the next year the following features are to be incorporated:

Neural network segmentation—using a neural network to determine which pixels are iris and which are non-iris in an iris image, based on local statistics.

ACKNOWLEDGEMENT

I am deeply indebted and would like to express my sincere thanks to our beloved Principal Dr. K.A.Krishnamurthy, for providing me an opportunity to do this project. My special gratitude to Dr. M.Z.Kurian, HOD, Department of E & C, S.S.I.T for his guidance, constant encouragement and wholehearted support. My sincere thanks to my guide Dr. M.Nagaraja, Asst. Prof, Department of E&C, S.S.I.T for his guidance, constant encouragement and wholehearted support.

REFERENCES

[1] Daugman, J., "Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression", IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 36, no. 7, July 1988, pp. 1169-1179.

[2] Daugman,J."How Iris Recognition Works", available at http://www.ncits.org/tc_home// m1htm/docs/m1020044.pdf.

[3] John Canny, "A Computational Approach to Edge Detection," IEEE Transactions of Pattern Analysis and Machine Intelligence, vol. PAMI- 8, no. 6, pp 679-698.

[4] J. Daugman, "Probing the uniqueness and randomness of IrisCodes: Results from 200 billion iris pair comparisons." Proceedings of the IEEE, vol. 94, no. 11, pp 1927-1935.

[5] Robert W. Ives, Randy Broussard, Lauren Kennell, Ryan Rakvic and Delores Etter, "Iris Recognition Using the Ridge Energy Direction (RED) Algorithm," Proceedings of the 42nd Annual Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, Nov. 2008.

[6] Ryan N. Rakvic, Bradley J. Ulis, Randy P. Broussard, and Robert W. Ives, "Iris Template Generation with Parallel Logic," Proceedings of the 42nd Annual Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, Nov. 2008.

[7] Altera, "TriMatrix Embedded Memory Blocks in Stratix IV Devices," Altera Corporation Stratix IV Device Handbook, vol 1, March 2010

◈◈◈

A Comparative Analysis with MVC Architectural Design And Implementation with Struts And Spring Framework

Tamal Dey

Dept. of MCA, PES Institute of Technology, 100 Feet Ring Road, BSK- III Stage, Bangalore- 560085
E-mail : tamal_dey@pes.edu

Abstract - Model-View-Controller architecture paradigm built over the integration of the components in Java framework and helps easy access of web services. MVC approach eases the horizontal development and maintenance of large scale distributed web applications in three frameworks. First is Big Blob framework, in this all the processing logics are part of GUI. Second is MVC architecture that easily works in command line or web interface. Third is Modified MVC Design. In implementation phase MVC architecture deals with J2EE and JSP-Servlet environment for web and internet programming. Also applying the MVC architecture with Multiple Frameworks gives privilege to work with few new concepts like Struts and spring components

Keywords - *User Interface, Modeling, and Visualization.*

I. INTRODUCTION

Web has the very complex issues these days. Since the desire of the companies and organizations are increasing so the complexity and the performance of the web programming matters. Complexity with the different types of communication devices is increasing. The business is demanding applications using the web and many communication devices. So with the increase load of the data on the internet we have to take care of the architecture issue [6].

Literature related to the topic is to identify Big Blog, MVC and Modified MVC architectural structures merits and demerits and make a set of comparison analysis for the architectural models in terms of implementation of each aspect. Next implement the MVC architecture with JSP-Servlet programming that helps to the architecture with multiple framework.

1. RELATED WORK

Various frameworks have been mentioned in literatures which are discussed below.

1.1 Big Blob Architecture

A common style of programming is to put all processing in the GUI. One common structure is "*big blob*" use this structure. It is useful for the absolutely simplest programs. This works correctly as long as the "model", the logic, is so small that it isn't worth putting into a separate class [2].

2.2 Model View Controller Architecture

Programming with graphical user interface (GUI) libraries makes easier to implement the model-view-controller (MVC) design. MVC was first introduced by Trygve Reenskaug, a Smalltalk developer at the Xerox Palo Alto Research Center in 1979, and helps to decouple data access and business logic from the manner in which it is displayed to the user [Figure 1]. More precisely, MVC can be broken down into three elements:

2.2.1 Model

The model represents data and the rules that govern access to and updates of this data. In enterprise software, a model often serves as a software approximation of a real-world process [2].

2.2.2 View

The view renders the contents of a model. It specifies exactly how the model data should be presented. When the model data changes, the view must update its presentation as needed. This can be achieved by using a *push model*, in which the view registers itself with the model for change notifications, or a *pull model*, in which the view is responsible for calling the model when it needs to retrieve the most current data [2].

2.2.3 Controller

The controller translates the user's interactions with the view into actions that the model will perform. In a stand-alone GUI client, user interactions could be button clicks or menu selections, whereas in an enterprise web application, they appear as GET and POST HTTP requests. Depending on the context, a controller may also select a new view -- for example, a web page of results -- to present back to the user [2].

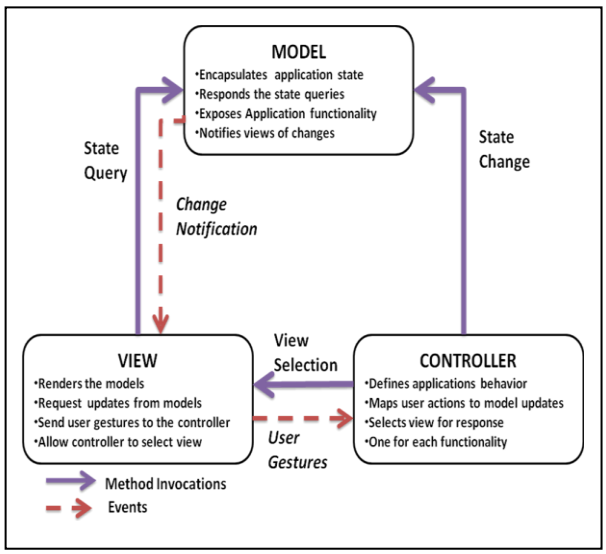


Fig 1: A Common MVC Implementation

2.3 Modified MVC Architecture

A more recent implementation of the MVC design places the controller between the model and the view. This design is common in the Apple Cocoa framework [Figure 2].

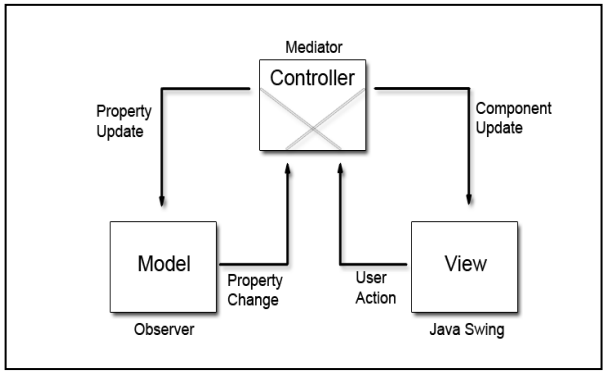


Fig 2: An MVC Design Placing the Controller between the Model and the View

The Cocoa and Cocoa Touch frameworks that power Mac OS X and iOS are tightly integrated into the Xcode development experience. Cocoa's high-level APIs make it easy to add animation, networking, and the native platform appearance and behavior to your application with only a few lines of code [6].

II. COMPARISION OF BIG-BLOB, MVC, MODIFIED MVC ARCHITECTURE

The *big blob* architectural programs are harder to read, maintain, and enhance. Paper can't really appreciate this when I start to build bigger programs. This works correctly as long as the "model", the logic, is so small that it isn't worth putting into a separate class. However, mixing model with presentation usually makes the program hard to read, and the inevitable growth of the program leads to a mess. This fails the simple Interface Independence test in web interface. Even changing model in any part also fails in this type of implementation [2].

The Model-View-Controller (MVC) architecture is easier to implement graphical user interface (GUI) libraries [2] because it provides a true decoupling of each part [3]. So it is easier to change the model part and view will notify the update the model does not carry a reference to the view but instead uses an event-notification model to notify interested parties of a change [Figure 3]. One of the consequences of this powerful design is that the many views can have the same underlying model. When a change in the data model occurs, each view is notified by a property change event and can update itself accordingly [3].

2.1 Solution with UI Model Architecture

Model View Controller structure can easily draw few possible solutions to implement the architecture. Firstly separate the user interface from the "model which makes huge improvement in simplicity, enhancements and maintenance are much easier with this structure [4]. Secondly model doesn't know about user interface. Lastly, model represent itself in text or graphically with the help of model class by override toString() method [5].

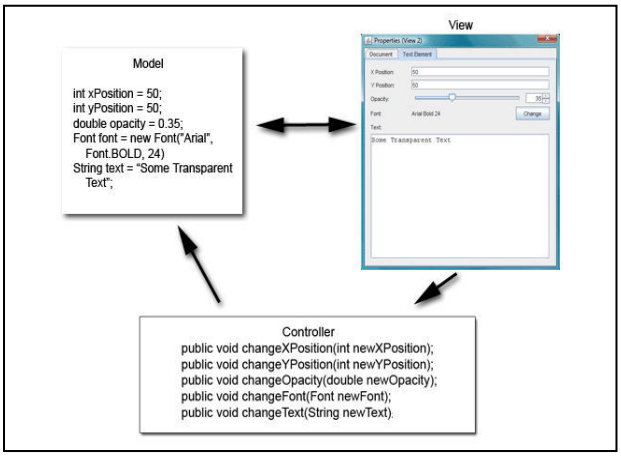


Fig 3: A Java SE Application Using MVC

The Modifying MVC architecture uses Cocoa Touch frameworks that power Mac OS X and iOS are tightly integrated into the Xcode development experience. Cocoa's high-level APIs make it easy to add animation, networking, and the native platform appearance and behavior to your application with only a few lines of code. In this Structure models encapsulate application data, Views display and edit that data, and Controllers mediate the logic between the two. By separating responsibilities in this manner, you end up with an application that is easier to design, implement, and maintain [3].

III. PROPOSED METHODOLOGY

This approach is based a combination of applying the two frameworks struts and spring for application development scenario. The sequence diagram for the combined application is explained, which is the main driving force for the application. This approach assumes that basic knowledge of web applications is essential. A testing has been done for the above concepts and found successful work. Major benefits of the above architecture are as follows.

1. It will provide a very clean division between actions like action forms, controllers, handlers, JavaBeans models, and views.

2. Spring's MVC is very flexible. Unlike Struts, this forces your Action and Form objects into concrete inheritance by using advantage of both.

3. Spring MVC is entirely based on interfaces. Every part of the Spring MVC framework is configurable.

4. It provides controllers, making it easy to handling of many requests from User Interface.

5. JSP or any other technology can be used to display the view, results to the user on the any of the output device.

6. Spring Controllers are configured via Inversion of Controls. This makes them easy to test and integrated with other objects managed by spring.

7. Spring MVC web tiers are typically easier to test as compared to Struts web tiers, due to the avoidance of forced concrete inheritance and explicit dependence of controllers on the dispatcher Servlet.

8. Struts framework was designed for the web interface purpose only. The spring framework was developed for the desktop and internet applications. When both frameworks used as combined it will provide the flexibility of implementation.

3.2 MVC architecture with JSP-Servlet environment

The MVC architecture was able to solve some of the problem of web and internet programming but still there were a lot of things missing from it. It was centred on the navigation of the JSP pages so there was the scope of the further development in the architecture point of view. During this process the next development was the Model 2 architecture. This problem was solved using the Servlet and JSP together. The Servlet handles the Initial request and partially process the data. It set up the beans then forward the result to the one of the JSP page. The Servlet decide the one of the page to be displayed from the list of pages [12].

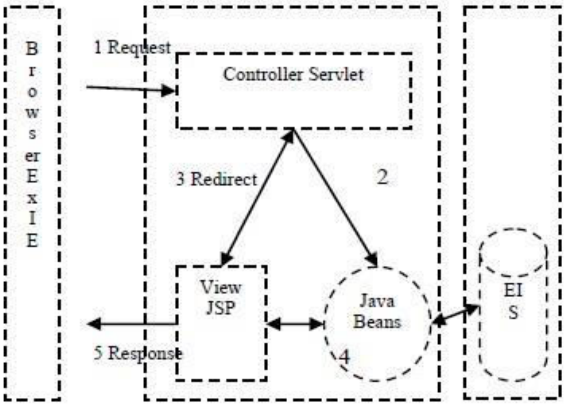


Fig 4: MVC architecture with JSP-Servlet

In this model all Control and application logic were handled by the Servlet. The Servlet was written in the java programming language. So it was also easier to handle the programming part of the Servlet. In this scenario the Servest becomes the power full for the complete application and it has emerged as the center point for the application. In the model architecture the Servlet becomes the gatekeeper for the all common tasks. It provides the common services like authentication, authorization, error control and follow of the application. This architecture has solved the most of the problems. But still there were many new issues emerged while applying this architecture [11].

3.3 Applying architecture with Multiple Frameworks

Web and Internet is ever growing area and the demands for the applications are growing. A single framework is not capable to handle the architecture of the application. To meet the currents requirement of the applications it's necessary to design a architecture to implement the frameworks [4]. Struts framework have been designed and developed for the front end control of the web applications [Figure 5]. It provides the various features for the applications that interact to the users. It also follows the MVC design features. Spring Framework is the designed to handle the various tasks.

The spring work for the desktop and internet based applications also. It follows the principals of the MVC. The simultaneous use of the Struts and spring frameworks in the single application with the applying the MVC Design principals so that can improve the performance of the applications. Struts Framework consists of three major blocks, Described in brief as follows [9].

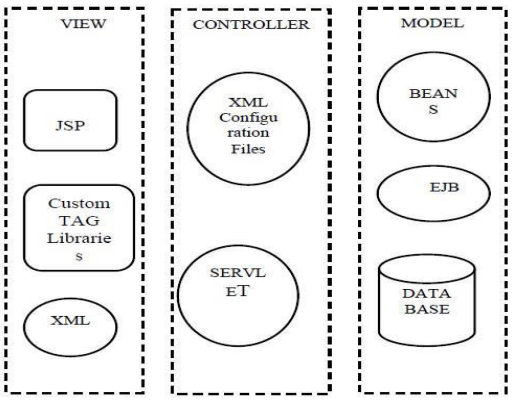


Fig 5: Struts model architecture

First is The View Block which controls the presentation part of the complete model. This contains following JSP files which you write for your specific application set of JSP custom tag libraries Resource files for internationalization [9].

Second Block is representing the Controller. This is for navigation the complete application. This contains XML configuration files; it contains the tags for the navigation of the paths [9].

Third Block is the Model. This part does the work of the Business Logic, Fetching and storing data to the database. This contains following Java Beans Enterprise Java Beans Database. Following figure shows the working of the components in the struts framework [9].

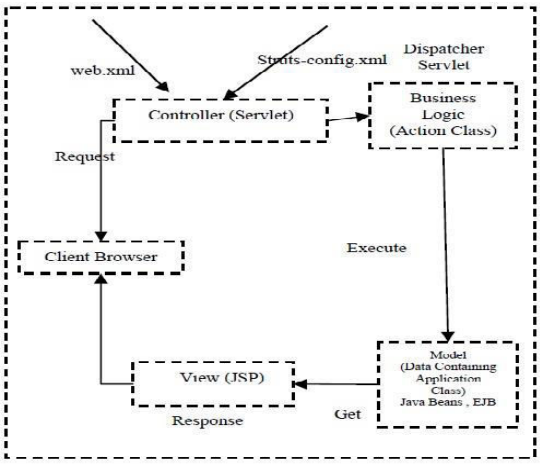


Fig 6: Component in Struts architecture

3.3.1 MVC in Struts

The major three parts of the MVC are as follows in the spring framework. Servlet controller (Controller Part) Java Server Pages or any other presentation technology (View Part) Application Business Logic: in the form of whatever suits the application (Model Part) [7].

Struts are a framework that implements a powerful and flexible controller based on the *Service To Worker* pattern. Struts' main advantages are: Integration flexibility: Struts' architecture provides flexibility for choosing the view and the model to be used. The view is based on the plug-ins concept. A plug-in is a dynamic mechanism by means of which a component or set of components that implement certain functionality in our application can be replaced by another ones, by simply modifying the application's configuration [Figure 7].

This model is implemented through JavaBeans, thus allowing its integration with other frameworks. It is supported by a solid community: Struts is a project from the Apache Software Foundation which has been consolidated as one of the most important organizations in the open source scope. In (Sing, 2002), SUN recommends using Struts as the framework for the Web tier [14].

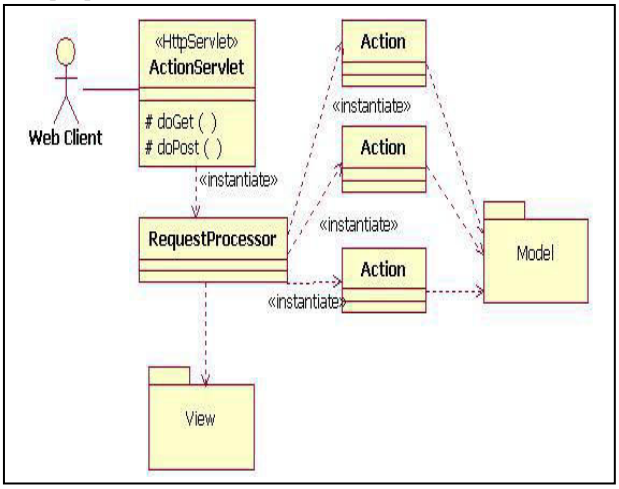


Fig 7: Struts's class diagram

Struts are composed of three main components:

1. The *ActionServlet* (the model's *Front Controller*), which is responsible for the application configuration and for receiving and analyzing the clients' requests. This component extracts from the configuration file (*strut-config.xml*) the general configuration parameters, the set of components that defines its behaviour (plug-ins) and the properties of each request. After performing these tasks, it delegates the control in the *RequestProcessor*.

2. The *RequestProcessor* (*Request Dispatcher* in the model), that creates an instance of the action (*Command* pattern) associated to the received request and executes it.

3. The *Action* (*Command* in the model). For eachoperation or use case, the developer creates an action (object) that inherits from the Action component. Each action is associated to a request type in Struts' configuration file.

3.3.2 Spring Components.

In the spring also follow the principals of the MVC. It has been designed more for the desktop and internet based applications. Spring consist of three core collaborating components [13].

1. Controller: Handles navigation logic and interacts with the Service tier for business logic.
2. Model: The contract between the Controller and the View Contains the data needed to render the View populated by the Controller.
3. View: Renders the response to the request Pulls data from the model. Core components in the spring MVC are as follows.

3.1 Dispatcher Servlet: Spring's Front Controller implementation.

It is the first controller which interacts to the requests that can also say it is an implementation of the Servlet. It controls the complete flow of the application.

3.2. Controller: User created component for handling requests encapsulates navigation logic delegates to the service objects for business logic.

3.3. View: Responsible for rendering output. Different views can be selected for the different types of output bases on the results and the viewing device, communication devices.

3.4. Model and View: It is the core part of the spring framework. It implements the business logic of the application. It is controlled by the controller. It stores the business logic and the view associated with it. Whenever it is executed it wills the data with the name of the view.

3.5. View Resolver: How the output is to be displayed depends on the result received from Model and View. It is used to map logical view names to actual view implementations. This part identifies and implement what is the output media and how to display it.

3.6. Handler Mapping: Strategy interface used by Dispatcher Servlet for mapping incoming requests to individual Controllers. It identifies the request and calls the respective handler to provide the services.

The following figure shows how the model will work. In this the dispatcher Servlet is the entry point for the application. The Struts parts do its work and send the request to the dispatcher Servlet [Figure 8]. The Servlet decides the handler. Then it will call to the controller. Controller will execute the Model and View [8].

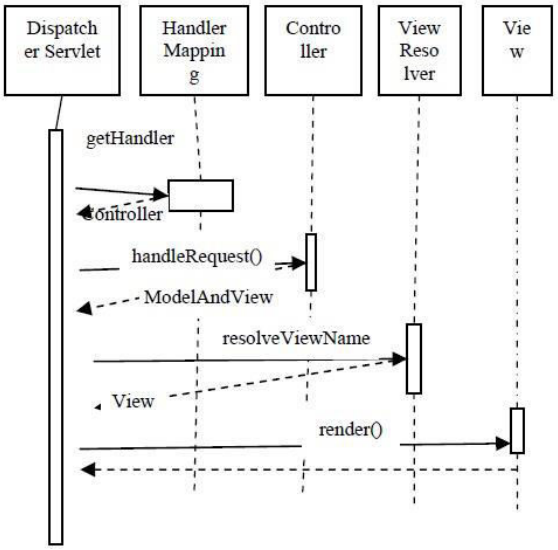


Fig 8: Sequence flow of application in the spring framework

IV. IMPLEMENTATION

Different frameworks based on Model that facilitate the development of J2EE applications. Some of these are integrated with servers and tools specific of their corresponding J2EE providers. There are also open source frameworks supported by a solid community and widely spread in the last years. In order to select the most appropriate framework for that purposes and carried out a survey which focused on open source frameworks, due to its can improve the performance of the Large Database application in terms handling number of requests. Inexpensive costs and the technological maturity reached by some of them. A suitable framework must achieve the following two objectives: first, it must adapt to our model's specifications, and second it must be a good framework – as described in the background. After analyzing the most used, widely spread open source frameworks in the Java community (Struts, Cocoon, Maverick, SOFIA, Spring, WebWork, Tapestry, Turbina and JSF), it observed that none of these completely satisfied the established requirements. So it has been decided to use a different framework for each of the model's parts (the Model, the View and the Controller): It chose Struts as the Controller, Cocoon for the View and StrutsEJB for the Model. In the next point, It describe each of these frameworks and how they fit into our model.

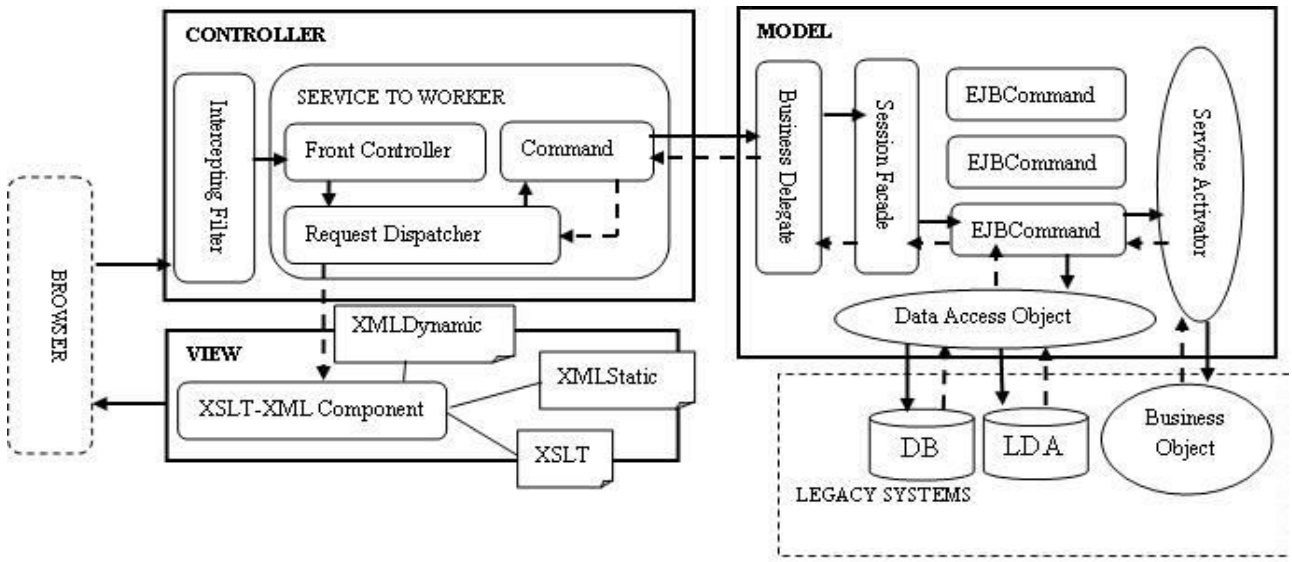


Fig 9: Model architecture

The primary target of the proposed model is to simplify the development of large applications based on the J2EE platform), thus providing a well structured architectural design, which allows for a complete decoupling of the system's main elements and synthesizes existing models, patterns and frameworks in the best way [14].

In this model, the controller serves as the application's entry point. It is implemented using only two patterns: the Intercepting Filter and the *Service To Worker*.

The *Intercepting Filter* is used in our model to implement the requests pre-processor; this system initially manages the entry requests from clients in the presentation layer. There are different types of requests, each one needing a particular processing scheme. Therefore, when a request arrives to the application, it should pass through a set of verifications before reaching the main processing phase – called the Front Controller – authentication, session validation, client IP address checking, request authorization, data codification, auditory or browser type used. The Intercepting Filter pattern is a flexible and highly decoupled way to intercept a request, applying a set of filters, thus rejecting or allowing the request to arrive to the initial process.

This initial process plays the controller's role: it analyzes each request to identify the operation to perform, thus invoking the business logic associated to each particular request and controlling the flow to the following view. In the proposed model [Figure: 9], our controller is designed following the Service To Worker pattern, which combines a set of smaller patterns that provide a complete and flexible solution to fulfil the requirements for an MVC controller while allowing the separation between actions – the model –, the view and the controller.

The Front Controller pattern describes a central point that manages the requests. In order to reach a greater flexibility and independence between the view and the model, the Front Controller assumes only the request analysis task, delegating in the Request Dispatcher the selection of the view and the action to perform. After the analysis phase, the Request Dispatcher will be in charge to select the command that encapsulates the operation to perform. Once this command has generated the result, the Request Dispatcher will select the next view to be shown to the user. Delegating these tasks in the Request Dispatcher gives our model a greater flexibility since it can introduce new views or models in the scenario by altering the component's behaviour.

The Command pattern represents each request by means of an object, therefore providing a very simple way to introduce new operations. In our model, the Command pattern is responsible for encapsulating the request information, parameters and the current state into a command object that contains the business logic. This command is then sent through the network to the model, where it is finally executed (EJB Command). By using this approach, it achieves a complete decoupling between the controller and the model, which represents the business data and implements the rules to operate

them. Following the same approach, in the model shown in [Figure 9], here introduce the business layer as a set of patterns that completely disconnect the controller and the view from the model, thus achieving MVC paradigm's objective. On the other hand, it defines another set of patterns in order to integrate our model with inherited systems or other business models.

In the first case it applys the Business Delegate, Session Facade and EJB Command patterns. The EJB Command pattern is a special case of the Command pattern where the business logic is encapsulated into a serializable object created by a remote client – the controller's Command – and sent through the network to the EJB container where it will be executed – by invoking its execute method. This scheme provides the advantages of the Command pattern in an environment where the business logic is distributed, therefore allowing the execution of business rules without overloading the application by a massive usage of EJBs.

For the second case, this has defined two patterns to ease the integration between the business model and the inherited systems or other business models. The Data Object Access pattern supplies a mechanism to abstract and encapsulate access to the data sources, therefore achieving warehouse independency. It also achieves a clear separation between the business logic and the data logic, increasing the applications' maintenance capabilities. The Service Activator pattern describes a way to access other business models and services in an asynchronous manner. When a message is received, The view is responsible for showing the data output by the MVC model. One of our models' goals is to decouple the presentation from the controller and the model, and to achieve this the model's output is first produced in XML format, for its later transformation by XSLT sheets into the final presentation shown to the client. XML/XSLT is an elegant way to separate the data from the presentation and to free it from any particular technology.

V. CONCLUSION

Multiple framework architecture works better as compare to any single framework architecture with the effective of the multiple frameworks for the development of the large scale applications. Hence, the controller mediates the flow of data between model and view objects in both directions in modified MVC architecture. View objects use the controller to translate user actions into property updates on the model. In addition, changes in model state are communicated to view objects through an application's controller objects. The state changes in model objects communicating to the view through the controller can be observed in more traditional version of MVC.

VI. FUTURE RESEARCH

Open technologies are the best to attract the academic and research scholar to work. J2EE is the vast field now a day, its open technologies also. Architecture is never fixed its goes on changing with the change in the technology. There are many frameworks available to work with J2EE technologies, Single frame is never sufficient to provide the complete solution with all essential features of the application. There is a lot of scope to work further with many other frameworks to implement and enhance the MVC architecture.

REFERENCES

[1] Apple Mac OS X as on 13th September 2011 http://developer.apple.com/technologies/mac/cocoa.html

[2] Java Notes as on 7th September 2011 http://leepoint.net/notes-java/index.html

[3] Java SE Application Design with MVC, Article by Robert Eckstein, March 2007

[4] Pattern Oriented Software Architecture- Vol. I by Frank Buschmann, Regine Munie, Hans Rohnert, Peter Sommerlad, Michal Stal - John Wiley & sons Publication, 2006

[5] Practical Object-Oriented Design with UML 2nd Edition by Mark Priestley, 2003

[6] Praveen Gupta et. al. / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 04, 2010, 1047-1051

[7] Erxiang Chen; Minghui Liu, "Research and Design on Library Management System Based on Struts and Hibernate Framework", in WASE International Conference on Information Engineering ICIE 09, 2009, Vol. 2, PP. 310-313

[8] Juanjuan Yan; Bo Chen; Xiu-e Gao, "Le Wang; Research of Structure Integration Based on Struts and Hibernate", in 2009 WRI World Congress on Computer Science and Information Engineering, 2009, vol. 7, PP. 530-534.

[9] Yonglei Tao: "Component- vs. application-level MVC architecture", in Frontiers in Education 2002 FIE 2002. 32nd Annual, 2002, Vol 1, PP. T2G-7 - T2G-10

[10] Meiyu Fang, "Design and Implement of a Web Examination System Using Struts and EJB", Seventh International Conference on in Web-based Learning 2008, 2008, pp. 25-28

[11] Shu-qiang Huang, Huan-ming Zhang," Research on Improved MVC Design Pattern Based on Struts and XSL", in Information Science and Engineering ISISE 08 International Symposium on, 2008, vol. 1 PP. 451 – 455.

[12] Wojciechowski, J.; Sakowicz, B.; Dura, K.; Napieralski, A.,"MVC model, struts framework and file upload issues in web applications based on J2EE platform", in Proceedings of the International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science 2004, PP 342-345.

[13] Wang Ning; Li Liming; Wang Yanzhang; Wang Yi-bing; Wang Jing, "Research on the Web Information System Development Platform Based on MVC Design Pattern",in IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008 , Vol 3,pp. 203-206

[14] Decoupling MVC: J2EE design pattern integration by Francisco macia-perez, virgilio gilart-iglesias, diego marcos-jorquera, juan manuel garcia-chamizo.

◈◈◈

Registration using Multi-Resolution Optimization for the Early Diagnosis of Alzheimer's Disease

Pournami P.N, Sithara Kanakaraj & V.K. Govindan

Dept. of Computer Science and Engineering, National Institute of Technology, Calicut, Kerala, India
E-mail : {pournamipn,sithara_mcs10,vkg}@nitc.ac.in

Abstract - Alzheimer's disease is an irreversible, progressive brain disease that leads to the death of brain cells. It affects the parts of the brain that control thought, memory and language. A complete cure for the disease has not been discovered yet. But detection at an early stage would however help to slow down the spread of the disease. A periodic Magnetic Resonance Imaging (MRI) scanning of the brain regions and its comparison would help in identifying the rate of degeneration of the brain cells and the intensity of treatment required to slow down the cell decay rate. The growth of the disease is scattered and dependent on each cell, so only an accurate and precise comparison result of the MRI would be helpful to identify the intensity of spread. In this paper, we present a technique to register the serial MRI images of a possible Alzheimer's disease subject. The growth of the disease can be analyzed by the image resulting from the difference of the registered image.

Keywords - *Alzheimer's Disease; Image registration; Multi-resolution; Serial MRI; Mutual information.*

I. INTRODUCTION

The application of computer related evaluation in the medical field has been growing day by day. The interpretations conveyed from the medical images help a lot in clinical diagnosis of the disease and about its growth/degradation. There are different kinds of modalities being used in medical imaging each one specifying a different use. Modalities like Computed Tomography (CT) and Magnetic resonance imaging (MRI) give details about the anatomical structures of the different body parts. Whereas, functional images like Single-photon emission computed tomography (SPECT) and Positron emission tomography (PET) represents the functional characteristics of that region. MRI gives a good contrast picture of the different soft tissues in a region and hence it proves to be a better imaging technique for brain images. The internal structures like the skull and the tissues inside can be differentiated very well in an MR image. In concern with brain disease, imaging is always done in a periodic nature to assess the growth of the disease in a more accurate way. A serial MRI would depict the progression in the state of the brain on a timeline basis with a gap of 6 months to 12 months or even more. The registration of serial MRI would make it much easier for identification of growth or loss of tissues in the brain.

In this paper, we have presented a method that will register the serial MRI of an Alzheimer disease (AD) subject to its baseline image. Upon registration, the rate of loss in the white matter of the subject's brain can be observed from which it is possible to predict the progressive status of the diseased brain. The paper is organized as follows. A review on the existing techniques for diagnosing Alzheimer's has been done in Section II. Section III describes the proposed algorithm for registration of an Alzheimer's disease person's brain. In Section IV the results and discussions related to the implemented algorithm has been portrayed. Finally, the outcome of the work is concluded in Section V.

II. RELATED WORKS

A brief review has been done in some of the important works in the areas of image registration [1,6], various techniques used for detection of Alzheimer's disease affected subjects [3,7,8,9,10,11,12,13, 14], some of the skull removal methods and segmentation [4,5] techniques.

The general methods used in medical image registration have been briefly outlined in [1]. The various issues faced during the registration process and its solutions along with an illustration of a few examples have been presented. The difficulty of the problem of image registration and the variability/ diversity of possible solutions has been discussed by the authors.

Linda K et al [2] studied whether the structural MRI can be used to identify a pattern of regional atrophy. Volumetric segmentation and cortical surface reconstruction methods were applied to the MRI from

mild AD and healthy control subjects. A stepwise linear discriminant analysis was done to identify regions that would best aid in discrimination of Healthy Control (HC) subjects from subjects with AD. Results showed a pattern of regional atrophy with high sensitivity and specificity which discriminated AD from HC subject. Thus the authors conclude that MRI can be used to identify patterns of regional atrophy.

A computer-based high dimensional pattern classification method that detects the entire spatial pattern of the brain structure was proposed by Christos Davatzikos et al. in [3]. The authors observed that, from the cross-sectional MRI scan, the complex and subtle structural patterns that characterize brain abnormalities in individual Mild Cognitive Impairment (MCI) could be identified. Also the individuals with MCI could be distinguished from cognitive healthy controls with high accuracy. Another classification framework based on MRI to distinguish AD patients from normal individuals has been presented in [4]. The method extracted the 3D volumetric features and 2D shape features from the MRI data and a Support Vector Machine (SVM) classifier was trained for classification of AD subjects. Using both, 2D and 3D features together, an accuracy of 84% was obtained when compared to using either one, 3D with 64% accuracy or 2D with 72% accuracy.

Olivier Querbes et al.[5] have observed a relation between the cognitive reserve, structural changes and the timeline of evolution to AD. Authors have concluded that the AD subjects would undergo structural changes in the brain which could be detected up to 24 months before the current clinical criteria. So a study on the serial MRI data of a subject could reveal the structural changes and diagnose it for a possible AD or not.

Based on Voxel-based morphometry, the white matter abnormalities in patients with mild AD and their topographic relationship with gray matter atrophy were studied in [6]. Results showed significant reduction in white matter regions of the AD group relative to controls. Also motor dysfunction was identified in moderate and even very mild AD. Another work using voxel-based morphometry was done by Michael et al. in [7] to determine locations of significant cortical atrophy in AD subjects. Authors observed that atrophy is prominent in the temporal cortex and frontal cortex, and also evident in the right hemisphere than the left hemisphere of the brain. Gold BT et al. in [8] have summarized the recent evidences suggesting that white matter declines are present in individuals at high AD-risk, prior to cognitive decline. From the results it is evident that in AD subjects a reduction in white matter occurs prominently in the temporal and frontal cortex that is associated with memory.

D Salas-Gonzalez et al. [9] proposed a computer-aided diagnosis technique for improving accuracy of early diagnosis of Alzheimer-type dementia on SPECT images. The methodology calculates mean of normal and mean of Alzheimer disease images using Welch's t-test. This method resulted in an accuracy of 96.2% in classification. However, the brain atrophy could not be measured in this method.

A non-rigid image registration technique was used by Mahanand and Aswatha [10] to analyze the brain tissue changes in AD. Initially the source and the target images were affinely registered and then non-rigid registration was performed. Results showed considerable deformation in the hippo-campus region during the screening and 6 month time interval. The changes seen in the brain tissues could help in tracking progress in treatment and possibly for identifying other neurodegenerative disease. More useful information regarding MCI and AD could be gathered if DTI or other such powerful analysis methods would be used along with MRI.

Shanthi and Sasi Kumar [11] proposed an automatic segmentation method for an MRI slice of the brain. Initially a T1 weighted MRI image was taken and skull stripping was done. Then using threshold value and seed growth technique the brain was segmented into White matter (WM), Gray matter (GM) and Cerebrospinal Fluid (CSF). Kyle and Ernest [12] have developed an accurate method for segmenting the MR brain images. But, the method is not fully automatic. The initial seed points are selected either using the threshold values or by user interaction. Using these seed points, region growing is performed in all the three dimensions. The algorithm claims to be fast and flexible.

Frederik Maes et al. [13] propose a method for multimodality medical image registration using mutual information. The intensities of the corresponding voxels in both the images are measured to find the statistical dependency between them. A maximum value indicates the correct geometrical alignment of the images. The results demonstrated sub-voxel accuracy without any segmentation or pre-processing steps.

The review carried out highlights on the pros and cons of some of the different techniques used as part of diagnosing AD. The survey reveals some facts on: loss of white matter is evident in AD patients; MRI slices can be used for identifying atrophy in regions of the brain; seed growing algorithm is a commonly used method for segmenting regions of interest; registration using mutual information has attained accuracy at sub-voxel level. The global registration method using mutual information has got a very high time complexity, which makes it unsuitable for real time applications. The problem addressed in this paper is to reduce the

computational time and also reduce the search space for registering the serial MRI slices.

III. PROPOSED METHOD

The registration process has played a crucial role in many applications related to the medical field where the changes in parts of the body have to be monitored frequently. The use of registration along with intensity based similarity metric, Mutual Information (MI), has developed to be an accurate measure for rigid and affine mono and multimodality image registration [14]. The global registration process using mutual information is however time consuming and therefore cannot be used in real time applications. The algorithm used in this paper has the following five major steps: Image preprocessing, Feature extraction, Similarity metric, Optimization and Image fusion. These are briefly presented in the following subsections. The raw images are initially pre-processed to make it applicable for the registration process.

A. Image Pre-processing

- Noise removal- The high frequency noise induced in the image can be effectively removed using filters. A low pass filter is used to remove the noise speckles caused by the high frequency signals.
- Skull Stripping- The MR image of a brain has both the skull and the brain tissues. The skull is not required in the process so it is removed. This process is called skull stripping. The image is divided into regions based on its perimeter. The region with the largest perimeter will be the skull in this case. This region is marked and then removed; resultant is an image without the skull.

B. Feature Extraction

The skull is removed and the image contains only the brain tissues. The brain tissue is a collection of three regions, namely, White Matter, Gray Matter and Cerebrospinal Fluid. To extract the White Matter from the brain tissue, we use seed region growing technique from [11]. A gray-level histogram of the image is plotted to identify the peak intensity of white matter and gray matter regions. These intensity values act as the seed. The resultant image contains the white matter of the MR slice.

C. Similarity Metric

Mutual information (MI) based on the Shannon definition of entropy is often employed for image registration [15]. MI explains how much one image is independent of the other image. A minimum value of MI shows total independence, and a maximum value shows that the respective transformation parameters results in the correct registration between the images. Registration using MI reduces the global spatial differences between corresponding images, which are caused by the positioning difference during their acquisition.

Entropy values are calculated from the histograms of the two images that are plotted based on the co-occurrence of their intensity values. The Shannon entropy is calculated using (1).

$$I(X;Y) = [H(X|Y) + H(Y|X)] - H(X,Y) \qquad (1)$$

Where H(X|Y) and H(Y|X) are the conditional entropies, and H(X,Y) is the joint entropy of X and Y.

High MI value indicates a lesser uncertainty while a lower MI value indicates a larger uncertainty. A zero MI value between the two random variables indicates their independent nature.

D. Optimization

In image registration, the search space is directly proportional to the runtime of the algorithm. So basically, the optimization strategy aims at reducing the search space. In this paper, Powell's multi-dimensional direction set method was combined with Brent's one dimensional method, to be used as the optimization strategy. Usually Powell's method is used to converge at minimum values, but in this algorithm, maximum value of MI is searched so as to acquire the maximal dependence between the two images. To optimize an N-dimensional data, initially a starting point P is selected. Then the method proceeds along any vector direction 'n' so as to find a maximum f (P), where 'f' is the similarity criteria, in that direction. Brent's one dimensional method is used for finding the maximum along each dimension 'n'. After finding the maximum P in one direction, that point is fixed as the maximum. This is repeated for all the n-dimensions and finally the algorithm converges at the maximum point where the MI gives the maximum value in all the dimensions. The maximum MI value results in the best registration alignment. Correspondingly a set of parameters, t_x, t_y, t_z (translations) and θ_x, θ_y, θ_z (rotations) are defined as the transformation that is required to align the two images.

To speed up the search in the 6 degrees-of-freedom (3 Rotations and 3 Translations) a multi-resolution technique using wavelet was incorporated into the algorithm. The wavelet method has two kinds of filters. One is a smoothing filter, which functions like a low pass filter and the other is a high pass filter. In the first level of decomposition, both the filters are applied to the image, this results in a high frequency coefficient and a low frequency coefficient whose resolution is half of the original data. The image is now in the frequency domain. The second level of decomposition is obtained

by applying the same procedure on the low frequency coefficients of the first level decomposition. The image is converted back to the spatial domain before any calculation is being done. At each of the decomposition stage the Powell's optimization is applied to reduce the search space. A course registration is done with the images of the second level decomposition. The resulting parameters of the transformation obtained, indicate the range in which the original image has to be transformed in order to approximately register with each other.

E. Image Fusion

The transformation values t_x, t_y, t_z and θ_x, θ_y, θ_z are applied on the input image so as to align it with the target image. The two images are fused using a graphical user interface (GUI) of the wavelet toolbox software called wavemenu in Matlab.

IV. RESULTS AND DISCUSSION

We applied our algorithm for the registration of a serial T2 weighted MRI image of an AD patient. The goal was to align the images of the same slice taken at successive time intervals so as to identify the rate of shrinkage occurred in the brain during the past years. The serial MRI data of the subject was taken for a period of 4 years with an interval of one year in between them. A progressing AD patient's brain images were used for the evaluation of the experiment. Four sets of DICOM images, each of size 256x256x32 were retrieved from the ADNI database [18]. The same slice of the brain from each of the four sets is selected for the algorithm assessment. The proposed algorithm is implemented on Matlab version R2010a on an Intel i5 core processor.

We start off with the pre-processing of slices to remove noise and to strip off the skull and gray matter leaving the result to be the white matter mass of the brain. These steps have been depicted in Fig. 1. Registration based on mutual information was used to register the segmented images to the baseline image. A transformation matrix with 6 degrees-of-freedom (3 Translations and 3 Rotations) was calculated. For optimizing the procedure Powell's multi-dimensional direction set method is incorporated into the algorithm and also a multi-resolution method using wavelet is added so as to reduce the computational time required for the algorithm and thereby making it more efficient for real-time applications.

A comparison of this algorithm with the global mutual information method is done here. The results are illustrated in Table 1. The correctness of the proposed method is validated using a simple and general approach. At first a known transformation is applied to an image and then using the proposed method the images are re-aligned. The applied transformation is taken as the ground truth [16, 17]. Table 2 shows results of some images that were used for verifying the correctness of the proposed approach. Fig. 2 shows the pre-processed slices of the serial MRI that are taken at a period of 4 years. Fig. 3 shows the difference in the consecutive slices of the serial MRI. The difference image shows the decay of white matter in the brain.

V. CONCLUSION

A full proof method for the early diagnosis of AD has not yet been discovered. This attempt using registration with multi-resolution has helped to visually diagnose the reduction in white matter mass of the brain. The work registers the MRI slice of human brain onto the baseline set and the difference image is calculated in-between each consecutive year's slice in order to visually detect the loss in the white matter. The algorithm exhibits a reduction in processing time when compared with the global registration method using mutual information. A reduction of one-fourth of the search space is also attained. The difference image got as the outcome of the subtraction of registered images depicts the growth of the disease.

TABLE I : COMPARISON RESULTS OF PROPOSED METHOD WITH THE GLOBAL REGISTRATION METHOD USING MUTUAL INFORMATION

IMAGE SET	**Existing**	**23**	**21**	**0**	**8.1°**	**0°**	**0°**	**589.9**
1	Proposed	23	21	0	8.1°	0°	0°	394.8
	Existing	24	24	0	-12°	0°	0°	846.6
2	Proposed	24	24	0	-12°	0°	0°	539.7
3	Existing	7	10	0	-4°	0°	0°	151.5
	Proposed	7	10	0	-4°	0°	0°	85.26
4	Existing	13	7	0	2.1°	0°	0°	151.3
	Proposed	13	7	0	2.1°	0°	0°	90.92

TABLE II : VALIDATION OF PROPOSED METHOD WITH GROUND TRUTH

No.	**GROUND TRUTH**						**EXPERIMENT RESULT**					
	Tx	***Ty***	***Tz***	***θx***	***θy***	***θz***	***Tx***	***Ty***	***Tz***	***θx***	***θy***	***Θz***
1	1	1	0	14.3°	0°	0°	1	1	0	14.3°	0°	0°
2	7	7	0	9.6°	0°	0°	7	7	0	9.6°	0°	0°
3	22	1	0	4.8°	0°	0°	22	1	0	4.8°	0°	0°
4	6	3	0	20.8°	0°	0°	6	3	0	20.8°	0°	0°
5	3	4	0	3°	0°	0°	3	4	0	3°	0°	0°

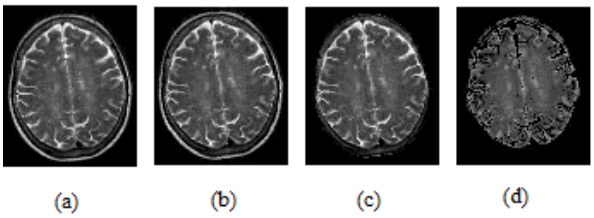


Fig. 1 : The above image depicts the pre-processing steps performed on each slice of the image. Fig1. (a) the original image, (b) De-noised image, (c) Skull Stripped and (d) white matter of the slice.

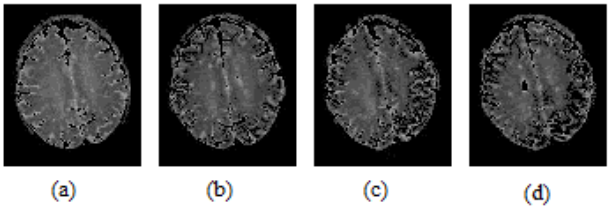


Fig. 2 : Resulting images of the pre-processed slices of (a) Baseline, (b) 12th month, (c) 24 th month and (d) 36 th month.

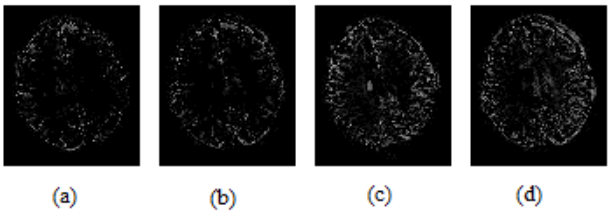


Fig. 3 : Difference image calculated between (a) Baseline and 12th month, (b) 12th month and 24th month, (c) 24th month and 36th month and (d) Baseline and 36th month

REFERENCES

[1] P.J. Kostelec and S. Periaswamy, "Image registration for mri," Modern Signal Processing, vol. 46, pp.161-184, 2003.

[2] L. K. McEvoy, C. F. Notestine, J. C. Roddey,D. J. Hagler, D. Holland, D. S. Karow,C. J. Pung, J. B. Brewer, A. M. Dale, "Alzheimer's Disease : Quantitative structural neuroimaging for detection and prediction of clinical and structural changes in MCI," Radiology, vol 251, April 2009.

[3] C. Davatzikos, Y. Fan, X. Wu, D. Shen, and S. M. Resnick, "Detection of prodromal Alzheimer's disease via pattern classification of MRI," Neurobiology of Aging, vol 29, Issue: 4, pp. 514-523, 2008

[4] J. D. Lee, S. C. Su, C. H. Huang, Wen-Chuin Xu, Y. Y. Wei, "Using volume features and shape features for Alzheimer's disease diagnosis," Fourth International Conference on Innovative Computing, Information and Control (ICICIC), 2009.

[5] O. Querbes, F. Aubry, J. Pariente, J. A. Lotterie, J. F¸ois De´monet,Ve´ronique Duret,Miche` le Puel, I. Berry, J. C. Fort, P. Celsis, "Early diagnosis of Alzheimer's disease using cortical thickness : Impact of cognitive reserve", Brain 2009, vol. 132, pp. 2036–2047.

[6] S. Xie, J. X. Xiao, G. L. Gong, Y. F. Zang, Y. H. Wang, H. K. Wu, and X. X. Jiang, "Voxel-based detection of white matter abnormalities in mild AD," Neurology, June . 2006 , vol. 66, pp.1845-1849.

[7] M. P. Lin, C. Devita, J. C. Gee, and M. Grossman, "Using voxel-based morphometry to examine atrophy behaviour correlates in Alzheimer's Disease and Fronto Temporal Dementia," Springer-Verlag, Berlin Heidelberg 2002.

[8] B. T. Gold, N. F. Johnson, D. K. Powell, C. D Smith, "White matter integrity and vulnerability to alzheimer's disease: preliminary findings and future detections," Biochim Biophys Acta., 2011.

[9] D Salas-Gonzalez, J M Go´rriz , J Ram´ırez , M Lo´pez, I A´ lvarez,F Segovia, R Chaves and C G Puntonet, "Computer-aided diagnosis of Alzheimer's disease using svm and classification trees," Phys. Med. Biol., vol. 55,pp. 2807–2817, 2010.

[10] B. S. Mahanand and M. A. Kumar, "Analysis of Alzheimer's disease progression in structural magnetic resonance images," WSEAS transactions on computers, vol. 8(4), April 2009.

[11] Shanthi K. J and M. Sasi Kumar, "Skull Stripping and Automatic Segmentation of Brain MRI Using Seed Growth and Threshold Techniques," International Conference on Intelligent and Advanced Systems, 2007.

[12] R. Kyle Justice and E. M. Stokely, "3-D Segmentation of MR Brain Images Using Seeded Region Growing," 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 1996.

[13] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality Image Registration by Maximization of Mutual Information," IEEE transactions on medical imaging, vol. 16(2), April 1997.

[14] J. P. W. Pluim, J. B. Antoine Maintz and M. A. Viergever, " Image registration by maximization of combined mutual information and gradient information," IEEE transactions on medical imaging, vol. 19(8), 2000.

[15] M. P. Wachowiak, R. Smolíková, G. D. Tourassi, and A. S. Elmaghraby, "Similarity metrics based on nonadditive entropies for 2D-3D multimodal biomedical image registration," Medical Imaging, Proceedings of SPIE, vol. 5032 (2003).

[16] F. P. M. Oliveira and J. M. R. D. Tavares, "Medical Image Registration," Computer Methods in Biomechanics and Biomedical Engineering, pp. 1-21 , 2012.

[17] E. D'Agostino, F. Maes, D. Vandermeulen, P. Suetens, "A viscous fluid model for multimodal non-rigid image registration using mutual information," Medical Image Analysis, vol. 7, pp. 565–575, 2007.

[18] Alzheimers Disease Neuroimaging Initiative, http://adni.loni.ucla.edu

◈◈◈

An Improvement to k-Means Spherical Clustering

S Mohammad Ghouse & E. Madhusudhana Reddy
Department of Computer Science & Engineering, Madanapalle Institute of Technology & Science,
Madanapalle, Chittor (dt), Andhara Pradesh, India
E-mail : mits.ghouse@gmail.com, e_mreddy@yahoo.com

***Abstract* -** Data clustering is an important data exploration technique with many applications in data mining. The k-means algorithm is well known for its efficiency in clustering large data sets. However, this algorithm is suitable for spherical shaped clusters of similar sizes and densities. The quality of the resulting clusters decreases when the data set contains spherical shaped with large variance in sizes. In this paper, we introduce a competent procedure to overcome this problem. The proposed method is based on shifting the center of the large cluster toward the small cluster, and recomputing the membership of small cluster points, the experimental results reveal that the proposed algorithm produces satisfactory results.

***Keywords* -** K-Means, Data Clustering, Cluster Analysis.

I. INTRODUCTION

The huge amount of data collected and stored in databases increases the need for effective analysis methods to use the information contained implicitly there. One of the primary data analysis tasks is cluster analysis, which helps the user to understand the natural grouping or structure in a dataset. Therefore, the development of improved clustering algorithms has been received much attention. The goal of a clustering algorithm is to group the objects of a database into a set of meaningful subclasses [3]. Clustering is the process of partitioning or grouping a given set of patterns into disjoint clusters. This is done such that patterns in the same cluster are alike, and patterns belonging to two different clusters are different. Clustering has been a widely studied problem in a variety of application domains including data mining and knowledge discovery [10], data compression and vector quantization [11], pattern recognition and pattern classification [7], neural networks, artificial intelligence, and statistics. Existing clustering algorithms can be broadly classified into hierarchical and partitioning clustering algorithms [17].

Hierarchical algorithms decompose a database *D* of *n* objects into several levels of nested partitioning (clustering), represented by a dendrogram, i.e., a tree that iteratively splits *D* into smaller subsets until each subset consists of only one object. There are two types of hierarchical algorithms; an agglomerative that builds the tree from the leaf nodes up, whereas a divisive builds the tree from the top down. Partitioning algorithms construct a single partition of a database *D* of *n* objects into a set of *k* clusters. Optimization based partitioning algorithms typically represent clusters by a prototype. Objects are assigned to the cluster represented by the most similar prototype. An iterative control strategy is used to optimize the whole clustering such that, the average squared distances of objects to its prototypes are minimized. These clustering algorithms are effective in determining a good clustering, if the clusters are of convex shape, similar size and density, and if their number *k* can be reasonably estimated. Depending on the kind of prototypes, one can distinguish k-means, k-modes and k-medoids algorithms. In k-means algorithm [8], the prototype, called the center; is the mean value of all objects belonging to a cluster. The k-modes algorithm [16] extends the k-means paradigm to categorical domains. For k-medoids algorithms [7], the prototype, called the "medoid"; is the most centrally located object in the cluster. The algorithm CLARANS, introduced in [20], is an improved k-medoids type algorithm restricting the huge search space by using two additional user-supplied parameters. It is significantly more efficient than the wellknown k-medoids algorithms PAM and CLARA, presented in [7].

Among clustering formulations that are based on minimizing a formal objective function, perhaps the most widely used and studied is k-means clustering. Given a set of *n* data points in real dimensional space, *Rd*, and an integer *k*, the problem is to determine a set of *k* points in *Rd*, called centers, so as to minimize the mean squared distance from each data point to its nearest center. Although the k-means method has a number of advantages over other data clustering techniques, it also has drawbacks; it converges often at a local optimum [2], the final result depends on the initial

starting centers. Many researchers introduce some methods to select good initial starting centers; you can see [5] and [6]. Other researchers try to find the best value for the parameter *k* that determines the number of clusters or the value of *k* must be supplied by the user. You can see [22] and [21]. In recent years, many improvements have been proposed and implemented in the K-means method; you can see [9]. The k-means clustering algorithm attempts to determine *k* partitions that optimize a certain criterion function. The average square error criterion, defined in (1), is the most commonly used (*mi* is the mean of cluster *Ci*, *n* is the number of objects in the dataset).

$$E=1/n \sum_{i=1}^{k} \quad \sum_{xeci}(x - m_i)^2 \qquad (1)$$

The average square-error is a good measure of the within cluster variation across all the partitions. Thus, the average square error clustering tries to make the *k* clusters as compact and separated as possible, and works well when clusters are compact clouds that are rather well separated from one another [12]. However, when there are large differences in the sizes or geometries of different clusters, as illustrated in Figure 1, the square-error method could split large clusters to minimize the square-error. The square-error is larger for the three separate clusters in left than for the three clusters in right, where the big cluster is split into three portions, one of which is merged with one of the two smaller clusters. The reduction in square-error (in right) is due to the fact that the slight reduction in square error due to splitting the large cluster is weighted by many data points in the large cluster. We propose a competent idea to solve this problem. There are very large number of clustering algorithms appeared [24], [12], [8], [23], [15], [1] and [25] but, in this paper, we focus on the k-means algorithm, a new procedure is added to the k-means algorithm makes it able to discover clusters with large variance in sizes with small separation between clusters.

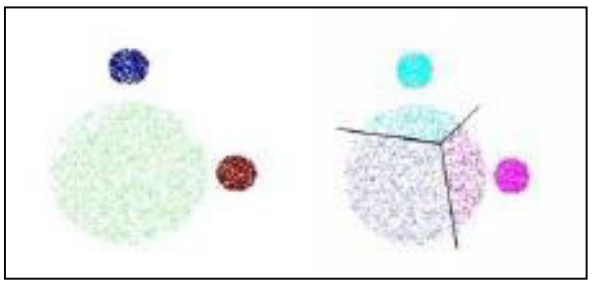


Fig. 1 Splitting of a large cluster by k-means algorithm.

II. RELATED WORK

The k-means algorithm uses the mean value of the objects in a cluster as the cluster center. Suppose that a dataset of *n* objects *x*1, *x*2 ,..., *xn* such that each object is in *Rd* , the problem of finding the minimum variance clustering of the dataset into *k* clusters is that of finding *k* points *mi*, *I* = 1, 2,..., *k* , in *Rd* such that Equation (1) is minimized. The basic processes of the k-means algorithm are:

1. Initialization: Select a set of *k* starting points *mj* , *j* = 1, 2,..., *k*. The selection may be done in random manner or according to some heuristic.
2. Distance calculation: For each object *xi*, 1≤ *I* ≤*n* compute its Euclidean distance to each cluster centroid *mj*, 1≤ *j* ≤*k*, and then find the closest cluster centroid.
3. Centroid recalculation: For each 1≤ *j* ≤*k* _ecomputed cluster centroid *mj* as the average of the data points assigned to it.
4. Convergence condition: Repeat step 2 and 3 until convergence.

Before the k-means algorithm converges, step2 and step3 are executed number of times, say *j*, where the positive integer *j* is known as the number of k-means iterations. The precise value of *j* varies depending on the initial starting clusters centroids even on the same data set. So the computational time complexity of the algorithm is O(*nkj*), Where *n* is the total number of objects in the dataset, *k* is the required number of clusters we identified and *j* is the number of iterations, *k* ≤ *n*, *j*≤ *n*. The k-means algorithm can be thought of as a gradient descent procedure which begins at the starting clusters centroids and iteratively updates these centroids to minimize the objective function in equation (1). It is known [4] that, the k-means will always converge to a local minimum. When we analyze the k-means we find that, the main advantages of this algorithm are; (1) its efficiency, (2) this algorithm is very easy to implement and (3) speed of convergence. On the other hand, its main drawbacks are (1) the final result depends on the initial starting centers, (2) to choose a proper number of clusters *k* is a domain dependent problem, (3) this algorithm is applicable only when mean is defined, (4) it is sensitive to outliers and (5) this algorithm is Good only for convex shaped, similar size and density clusters. For the first four disadvantages, there are a lot of efforts have been done to overcome these problems, we review some of them, the proposed method handles the last problem.

Several variants of the k-means algorithm have been proposed. Their purpose is to improve efficiency or find better clusters; improved efficiency is usually accomplished by either reducing the number of iterations to reach final convergence or reducing the total number of distance calculations. Therefore, choosing a good set of initial cluster centers is very important for the algorithm. However, it is difficult to

select a good set of initial cluster centers randomly. Bradley and Fayyad [5] have proposed an algorithm for refining the initial cluster centers. The main idea of their algorithm is to select *m* subsamples from the data set, apply the k-means on each subsample independently, keep the final *k* centers from each subsample provided that empty clusters are not be allowed, so they obtain a set contains *mk* points. They apply the k-means on this set *m* times; at the first time, the first *k* points are the initial centers. At the second time, the second *k* points are the initial centers, and so on. And the algorithm returns the best *k* centers from this set. They use 10 subsamples from the data set, each of size 1% of the full dataset size. To choose a proper number of clusters *k* is a domain dependent problem. To resolve this problem, some methods have been proposed to perform k-clustering for various numbers of clusters and employ certain criteria for selecting the most suitable value of *k* [21] and [22]. For example in [22] the authors depend on the fact that, the k-means method aims to minimize the sum of squared distances from all points to their cluster centers, this should result in compact clusters. So they use the distances of the points from their cluster centers to determine whether the clusters are compact or not. For this purpose, they use the intra-cluster distance measure, which is simply the distance between a point and its cluster center and take the average of all of these distances, as defined in equation (1). So the intra-cluster distance is the average squared error that the k-means method minimizes. Also the authors measure the inter-cluster distance, or the distance between clusters, which should to be as large as possible. So they calculate this as the distance between cluster centers, and take the minimum of these values, defined as in equation (2)

$$\text{Int er} = \min(\| m_i - m_j \|^2) \qquad (2)$$

$$i = 1,2,...,k \sqcup 1 \text{ and } j = i \sqcup 1,...,k$$

They take only the minimum of these values as they want the smallest of this distance to be maximized, use these measures to help them to determine if they have a good clustering, so they minimize the ratio between them, defined as in equation (3).

$$\text{Validty} = \frac{int\ ra}{int\ er} \qquad (3)$$

Therefore, the clustering which gives a minimum value for the validity measure will give the ideal value of *k* in the k-means algorithm. The k-means algorithm is applicable only when mean is defined, also this problem is solved by introducing the k-modes algorithm [16]. This is an extended version of the k-means with some modification to be suitable for categorical data. The cause that the k-means algorithm cannot cluster categorical objects is its dissimilarity measure and the method used to solve the clustering problem. These barriers have been removed by making the following modifications to the k-means algorithm

1. Using a simple matching dissimilarity measure for categorical objects.
2. Replacing means of clusters by modes.
3. Using a frequency-based method to find the modes to solve the problem.

It is known that the k-means is sensitive to outliers, but there are some researchers have been done to solve this problem. Some of them are to detect the outliers [14] first and remove them, and then apply a clustering algorithm. In [13] an Outlier Removal Clustering (ORC) algorithm is proposed, that provides outlier detection and data clustering simultaneously. This algorithm consists of two consecutive stages, which are repeated several times. In the first stage, they perform K-means algorithm with multiple initial starting points, and pick the best centers, and in the second stage, the algorithm assigns an outlyingness factor for each point and iteratively removes the points which are far from their clusters centers. This factor depends on the point's distance from the cluster center and the most far point from the cluster center. And all points with factor greater than specified threshold are considered as outliers and removed from the data set. The k-means algorithm is good only for convex shaped, similar sizes and density clusters. We propose a competent method in the following section to overcome this problem.

III. REFINEMENT OF THE FINAL CLUSTERS

The k-means algorithm is a popular clustering algorithm and has its application in data mining, image segmentation [22], bioinformatics and many other fields. This algorithm works well with spherical shaped clusters of similar sizes. In this section we present how to make this algorithm works well with spherical shaped clusters of any size. In our proposed method, we find the distances between the means result from the k-means algorithm. For each cluster, we calculate its average radius of a cluster *Ci* -by dividing the sum of squared error of its points from its representative by the number of points assigned to it- as in equation (4).

$$\text{Radius}(C_{i)} = \frac{\sum_{peci} d^2(p,m_i)}{n_i} \qquad (4)$$

We search for the largest cluster (have the largest average radius) and test whether this cluster have some portion merged with other clusters or not. At the first time, you can say if the summation of the two radiuses is larger than the distance between the two clusters, then there is a portion of the large cluster merged with the other cluster. So we can redistribute the points in the smaller cluster only over the two clusters. But this

formula is not suitable at all. Since the mean of cluster is the center of gravity of the points. Also the average radius (as in equation 4) for both clusters is larger than the actual radius, and the small cluster attracts portion of points from the largecluster, and this leads to enlargement of the actual radius of the small cluster, and the mean of the smaller cluster is attracted toward the larger cluster, since the objective of the kmeans is to get the smallest value for the squared error function in equation (1). So, we use equation (6) to get the sum of the two radiuses and compare this value with the distance between the two clusters which calculated as in equation (5). Where *mL* and *ms* are the means of the large and the small cluster respectively, *d* is the dimensionality of the data, *L* and *S* refer to the large and small cluster respectively.

$$\text{MeadnDistance} = \sqrt{\sum_{i=1}^{d}(m_{li} - m_{si})^2} \qquad (5)$$

$$Somofradius=(radius(L)+radius(S))*0.8 \qquad (6)$$

If the *Sumofradius* is larger than or equal to the *Meandistance* and at the same time the ratio between the two radius is smaller than 0.90, -this condition is used to insure that there is large difference in size-, then some portion of the larger cluster is merged with the smaller one. In this case we must redistribute the points in the smaller cluster to return the misclassified points to the larger cluster. How can we redistribute the points in the smaller cluster? To do this operation, we take the average of the two means as new mean. This new mean is located in the large cluster, and at the mid distance between the two cluster centers. We redistribute the points in the small cluster over the new mean (*Avmean*) and it's original mean (*ms*). Since the means of the two clusters are shifted. So, we use the next formula to redistribute the points of the smaller cluster. We add a small value to the right hand side of relation 7, since we expect a small separation between the two clusters and the mean of the smaller cluster is attracted to the larger cluster. We multiply the ratio between the two radiuses by 0.80 since the radius of the smaller cluster has error percent larger thanthe other cluster

$$\text{Dis(pi, Avmean)} \leq \text{Dis}(p_i, m_s) + \frac{radius(L)*0.8}{radius(s)}$$

If the formula 7 is true then the point *pi* is moved to the larger cluster, otherwise it remains in the smaller cluster. After redistribution of the all points in the smaller cluster, we recalculate the new means for the two clusters. All these processes are repeated for all clusters. So the final means of the proposed method are better than those produced first from the k-means algorithm. Fig.2 shows the proposed function added to the k-means algorithm to improve the final results.

Refinement process of the final results of the k-means algorithm

```
For i=1 to k
    For j=i+1 to k
        Compute Euclidean distance d(m1,m2)
    Next f
Next f
For j-1 to k/2
    Find the largest Cluster L
For i=1 to k
If(radius(L)>radius(I)
S=I
  Sum_of_radius = (radiys(L)+radius(S)*0.80)
        If(Sum_of_rsdius>=d(m1,m2)&&(radius(S)/rad
        ius(L))<0.9)
                Av_mean = (m1+m2)/2
                Redistribute points represented
                by m1 over the twp means m2 and
                Av_mean.
                All points assigned to
                Av_mean are moved to the cluster m2
                Find the new means for the
                cluster mt
        Endif
Endif
Next i
        Find the new mean for the cluster mt
Next j
```

IV. EXPERIMENTAL RESULTS

In this section, we present some experimental evaluation of the proposed algorithm, which reveal a great improvement in the k-means algorithm when the dataset contains spherical shaped clusters with large variance in their sizes. We have created many 2-dimensional datasets that contain spherical clusters of large different size. These datasets are created according to the general equation of circle $(x-a)^2 + (y-b)^2 \leq r^2$. We present here some of them. All these datasets contain three clusters as shown in Figures from 3 to 6. Table below presents the exact number of points in each cluster. Also we present the exact number of points in each cluster founded by the k-means and by our proposed algorithm. The experimental results in Table

show the great improvement at the final clusters discovered by the proposed algorithm.

COMPARISON BETWEEN THE RESULTS FROM THE K-MEANS AND OUR PROPOSED METHOD

Data set	Exact clusters	k-mean clusters	k-means error	Proposed clusters
Set1	1815	1210	605 points	1815
	683	973		683
	660	975		660
Set2	1582	1025	557 points	1582
	703	1015		703
	642	887		642
Set3	1582	1043	539 points	1585
	557	816		552
	522	802		524
Set4	2129	1306	823 points	2067
	505	916		534
	510	922		543
Set5	2363	1417	946 points	1417
	522	992		992
	524	1000		1000

When we examine the percent of error in the proposed algorithm, we find that, the proposed method produces the exact clusters in dataset 1 and dataset 2, because there is a separation between clusters. But there is a small error at clusters discovered from dataset 3, note that, there is no separation between clusters. From Table below you can see that, there are 7 points misclassified.

In this paper, we have described a new procedure added to the end of the k-means clustering algorithm. The objective of this procedure is to refine the results of the k-means. This procedure is optional and it is strongly recommended to use it after the k-means especially when the dataset contains spherical shaped clusters with large difference in their sizes. Our experimental results are evidence that our proposed method improve the quality of the resulting clusters. Our proposed algorithm produce the same result as k-means when the centers of the smaller clusters lie out of them, because in this situation the clusters seem to have very small difference between their radius. In future work we will search for more robust solution for this problem.

ACKNOWLEDGMENT

This is the project done by me in our collage as regular curriculum of M tech degree. This project was funded by MITS Collage of engineering.

REFERENCES

[1] R. Agrawal , J. Gehrke , D. Gunopulos, and P. Raghavan, "Automatic SubspaceClustering of High Dimensional Data for Data Mining Applications," Proc.ACM SIGMOD Int. Conf. on Management of Data, Seattle, WA, 1998, pp. 94-105.

[2] M R. Anderberg Cluster Analysis for Applications, Academic Press, 1973.

[3] M. Ankerst , M. Breunig , H.P. Kriegel , and J. Sander, "OPTICS:Ordering Points to Identify the Clustering Structure," Proc. ACM SIGMOD Int. Con. Management of Data Mining, 1999, pp. 49-60.

[4] L. Bottou, and Y. Bengio, "Convergence properties of the k-mean algorithm," The MIT press, Cambridge, MA, 1995, pp. 585-592. [5] P. S. Bradley , and U. M. Fayyad, "Refining Initial Points for K-Means Clustering," Proc. of the 15th International Conference on Machine Learning (ICML98), J. Shavlik (ed.). Morgan Kaufmann, San Francisco, 1998, pp. 91-99.

[5] S. Deelers, and S. Auwatanamongkol, "Enhancing K-Means Algorithm with Initial Cluster Centers Derived from Data Partitioning along the Data Axis with the Highest Variance," PWASET vol 26, 2007, pp. 323-328.

[6] R.O. Duda , and P.E. Hart, Pattern Classification and Scene Analysis. John Wiley Sons, New York, 1973.

[7] M. Ester, H.P. Kriegel, J. Sander, and X. Xu, "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining,AAAI Press, Portland, 1996, pp.226-231.

[8] A. M. Fahim, A. M. Salem ,F. A. Torkey, and M. Ramadan, "An efficient enhanced k-means clustering algorithm," Journal of Zhejiang University SCIENCE A, vol 7(10), 2006, pp. 1626-1633.

[9] U.M. Fayyad , G. Piatetsky-Shapiro , P. Smyth , and R. Uthurusamy Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1996.

[10] A. Gersho, and R.M. Gray Vector Quantization and Signal Compression, Kluwer Academic, Boston, 1992.

[11] S. Guha , R. Rastogi, and K. Shim, "CURE: An Efficient Clustering Algorithms for Large Databases," Proc. ACM SIGMOD Int. Conf. on Management of Data, Seattle, WA, 1998, pp. 73-84.

[12] V. Hautamaeki , S. Cherednichenko , I. Kaerkkaeinen , T. Kinnunen, and P. Fraenti, "Improving K-Means by Outlier Removal," SCIA 2005,LNCS 3540, 2005, pp. 978-987.

[13] V. Hautamaeki , I. Kaerkkaeinen, and P. Fraenti, "Outlier detection using k-nearest neighbourgraph," In: 17th International Conference on Pattern Recognition (ICPR 2004), Cambridge, United Kingdom, 2004, pp. 430-433.

[14] A. Hinneburg, and D. Keim, "An Efficient Approach to Clustering in Large Multimedia Databases with Noise," Proc. 4th Int. Conf. on Knowledge Discovery and Data Mining, New York City, NY. 1998.

[15] Z. Huang, "A fast clustering algorithm to cluster very large categorical data sets in data mining," Proceedings of the SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, Dept. of Computer Science, The University of British Columbia, Canada, 1997 pp. 1-8.

[16] A. K. Jain, and R. C. Dubes, Algorithms for Clustering Data, Prentice Hall, 1988.

[17] L. Kaufman, and P. Rousseeuw, "Finding Groups in Data: An Introduction to Cluster Analysis," Wiley, 1990.

[18] J.B. MacQueen, "Some methods for classification and analysis of multivariate observations. Proc. 5th Symp. Mathematical Statistics and Probability, Berkelely, CA, Vol(1), 1967, pp. 281297.

[19] R.T. Ng, and J. Han, "Efficient and Effective Clustering Methods for Spatial Data Mining," Proc. 20th Int. Conf. on Very Large Data Bases. Morgan Kaufmann Publishers, San Francisco, CA, 1994, pp. 144-155.

[20] D. T. Pham , S. S. Dimov, and C. D. Nguyen, "Selection of k in K-means clustering," Mechanical Engineering Science, vol(219), 2004, pp. 103-119.

[21] S. Ray, and R. H. Turi, "Determination of number of clusters in k-means clustering and application in colour image segmentation," In Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques, 1999, pp.137-143.

[22] G. Sheikholeslami, S. Chatterjee, and A. Zhang, "Wave-Cluster: A Multi-Resolution Clustering Approach for Very Large Spatial Databases," Proc. 24th Int. Conf. on Very Large Data Bases. New York,1998, pp. 428-439.

[23] R. Sibson, "SLINK: an optimally efficient algorithm for the single-link cluster method," The Comp. Journal, 16(1), 1973, pp. 30-34.

[24] T. Zhang, R. Ramakrishnan, and M. Linvy, "BIRCH: An Efficient Data Clustering Method for Very Large Databases," Proc. ACM SIGMOD Int. Conf. on Management of Data. ACM Press, New York, 1996, pp.103-114.

❑❑❑

Factors Influencing Job Rejections in Cloud Environment

Rashmi K S, Suma. V & Vaidehi. M

Dayananda Sagar College of Engineering, Bangalore, India
E-mail : rashmiks.ks57@gmail.com , sumavdsce@gmail.com, dm.vaidehi@gmail.com

Abstract - The IT organizations invests heavy capital by consuming large scale infrastructure and advanced operating platforms. The advances in technology has resulted in emergence of cloud computing, which is promising technology to achieve the aforementioned objective. At the peak hours, the jobs arriving to the cloud system are normally high demanding efficient execution and dispatch. An observation that has been carried out in this paper by capturing a job arriving pattern from a monitoring system explains that most of the jobs get rejected because of lack of efficient technology. The job rejections can be controlled by certain factors such as job scheduling and load balancing. Therefore, in this paper the efficiency of Round Robin (RR) scheduling strategy used for job scheduling and Shortest Job First Scheduling (SJFS) technique used for load balancing in reducing the job rejections are analyzed. Further, a proposal for an effective load balancing approach to avoid deadlocks has been discussed.

Keywords - *Round Robin Scheduling technique, Deadlock Avoidance, Migration, Load Balancing, Task scheduling, Shortest Job First Scheduling, Cloud Computing.*

I. INTRODUCTION

The IT organizations are oriented towards the development of cost, time and resource effective products. Cloud computing is one of the promising technology to achieve the aforementioned objective. Further, cloud computing is a pay-go-model which provides on demand network access to a shared pool of computing resources that can be rapidly provisioned and released with minimal management to its clients as a metered service. But in an IT organization jobs arriving to the system are normally high at peak hours, which demands efficient execution. Further, virtualization is the core characteristic of cloud computing, where the jobs that arrive to the system are diverse in nature and runs independently on the virtualized operating system. Consequently, to satisfy the request, the resources have to be properly utilized such that there is no deterioration in the working of the system which in turn leads to high rejection of the jobs. If the job rejection rate is high then the business of the cloud service provider also deteriorates. As cloud is a pay-go-model, the business performance needs to be accelerated which is a challenging issue in the domain. There are certain important factors namely job scheduling, load balancing and resource allocation in the cloud environment that supports to improve the business performance by reducing the job rejections.

An efficient job scheduling strategy must aim to yield less response time so that the execution of submitted jobs takes place within a stipulated time and simultaneously there will be an occurrence of in-time resource reallocation. As a result of this, less rejection of jobs takes place and more number of jobs can be submitted to the cloud by the clients which ultimately results in accelerating the business performance of the cloud system.

As the resources are provided on demand in the cloud system there is a necessary for the resources to be highly available. Failure of which, results in job starvation, which is a situation where a job does not get the resource that it requires for a long time because the resource is being allocated to other jobs. Hence, resource management is a complex task from the business point of the cloud service provider. Henceforth, an effective load balancing strategy must provide a solution to the aforementioned issue by yielding less response time which in turn results in less starvation and reduced job rejections which are the necessary factors to accelerate the business of the cloud service provider. However, in the cloud environment there are possibilities of deadlock occurrences when there is more than one request arriving to the system for competing to acquire the same resource at the same instance of time. Therefore, an effective load balancing strategy must be incorporated to avoid deadlocks.

Henceforth, in this paper a survey on the factors that governs to accelerate the business performance of the cloud has been provided.

The paper is organized as follows. The Section II of this paper explains the related work, Section III elucidates the research design and Section IV explains the research methodology. Finally, the paper is concluded in Section V followed by references.

II. RELATED WORK

The emergence of cloud computing in the IT industry has opened several research avenues in the domain. Authors of [1] have proposed an improved activity-based costing [IABC] algorithm for task scheduling in cloud computing with the objective to schedule group of tasks in cloud computing platform with resources having different costs and computation performance. The result of the algorithm shows that the time taken to complete the tasks is less by grouping them than without grouping them. However, this algorithm cannot handle simultaneous tasks.

However, authors in [2] have suggested a task scheduling technique using credit based assignment problem in cloud computing to find minimal completion time of tasks. They have considered a single parameter i.e., cost to minimize the completion time of tasks. However, in the cloud environment it is necessary to achieve optimal solution. Henceforth, it is imperative to consider all parameters which influence the realization of minimal completion time of tasks.

Authors in [3] have implemented efficient resource provisioning in compute clouds through VM multiplexing. In this paper, resources provisioning is based on the estimate of the capacity needs of the virtual machines. Further, the virtual machines are multiplexed using VM selection method, which uses the correlation among the VMs as an indicator of their compatibility. Hence, with this approach high utilization of the resources is achieved. But, the limitation of this approach is that there is an overhead involved during the VM selection process.

Authors in [4] have proposed a load balancing strategy for virtual storage to provide a large scale net data storage model and Storage as a Service model based on cloud. Further the storage virtualization is achieved using three layers architecture with two load balancing modules to balance the load. The strategy implemented in this paper is limited to the cloud service providers providing Storage as a Service (SaaS).

Authors in [5] have recommended load balancing in a three–level cloud computing network, by using a scheduling algorithm which combines the features of Opportunistic Load Balancing (OLB) and Load Balance Min-Min (LBMM) which can utilize better executing efficiency and maintain load balancing of the system. The objective is to select a node based for executing the complicated tasks that needs large-scale computation. The scheduling algorithm proposed in this paper is not dynamic and also there is an overhead involved in the selection of the node.

III. RESEARCH DESIGN

An effective resource model is very much necessary as it impacts the performance of the cloud computing [14]. As cloud is a pay-go model, it is necessary to have an efficient task scheduling strategy and an effective load balancing strategy to accelerate the business performance of the cloud system. Hence, it is essential that the resource model should support optimized utilization of available resources and execution of tasks within a stipulated time in the cloud environment.

In this investigation, a secondary data has been collected for analyzing the efficiency of the cloud service provider for in time execution of jobs at the peak hours with available resources. The secondary data is the processed data collected from several IT industries functioning in the cloud environment.

It is apparent from the analysis that task scheduling and load balancing strategies are the influencing factors in cloud to reduce job rejections and there by accelerates the business performance of the cloud system.

IV. RESEARCH METHOD

Figure 1. is the flow diagram of this investigation depicting the job processing in cloud . In this diagram diverse requests arrive from various users to the cloud datacenter. The cloud datacenter process these requests and dispatches them. The processing of request will be either through the job scheduling or through the load balancing. The existing job scheduling strategy operates in Round Robin (RR) fashion to yield less response time for reducing the job rejections at the peak hours while Shortest Job First Scheduling (SJFS) technique is used for load balancing, also yields less response time to reduce job rejections as well as to reduce starvations. Further, the load balancing should avoid deadlocks by promoting Virtual Machine (VM) migrations which is another important factor to reduce the job rejections. Subsequently, the reduction in job rejection ultimately results in accelerating the business performance of the cloud system

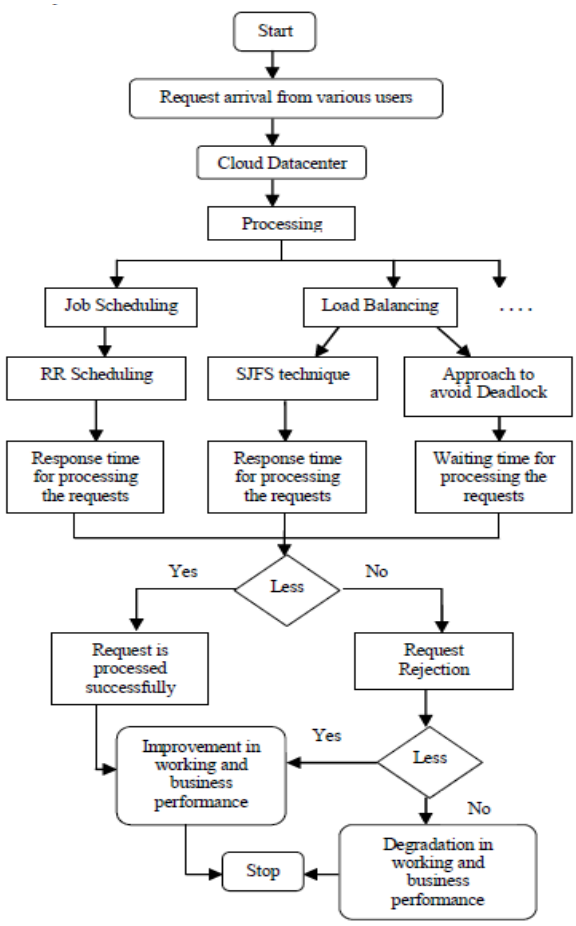


Fig. 1 : Job Processing in Cloud

The existing job scheduling technique in cloud is the Round Robin (RR) scheduling technique. Henceforth, in order to analyze the efficiency of RR scheduling technique a simulation is carried out using CloudAnalyst simulator. TABLE I., TABLE II. and TABLE III. depicts the simulation configuration.

TABLE I. depicts the detalis of user base configuration, which provides details related to request for every hour, data size for every request and datacenter where the requests are processed.

TABLE I : USER BASE CONFIGURATION

UB	R/U/H	DS/R	DC
UB1	12	100	DC4
UB2	12	10000	DC3
UB3	12	100000	DC1
UB4	12	1000	DC2
UB5	12	10000	DC2

UB – User base, R/U/H –Request per user per hour, DS/R – Data Size per request, DC – Datacenter.

TABLE II. shows the Datacenter configuration, which provides details related to the number of VMs present in DC, memory and bandwidth capability of every VM.

TABLE II : DATACENTER CONFIGURATION

DC	No. of VMs	Mem.	BW
DC1	40	1024	1000
DC2	20	512	100
DC3	50	512	10000
DC4	35	1024	1000

DC – Datacenter, Mem – Memory, BW - Bandwidth

TABLE III. shows the information of advanced configuration, which comprises of user grouping factor in user base, request grouping factor in datacenters and executable instruction length for every request.

TABLE III : ADVANCED CONFIGURATION

User grouping factor	1000
Request grouping factor	100
Executable instruction length	250

The result of this simulation has been obtained in terms of response time and processing time and TABLE IV. depicts the result of this simulation.

TABLE IV : RESPONSE TIME AND PROCESSING TIME OBTAINED USING RR SCHEDULER

Mm	RT (ms)	DC PT (ms)
Avg. (ms)	115.23	2.51
Min. (ms)	44.33	0.08
Max. (ms)	380.03	7.98

Mm – Measurement metric, RT –Response time, DC PT – Datacenter Processing Time

Figure 2. shows a sample graph of user base hourly average response time using the RR scheduler.

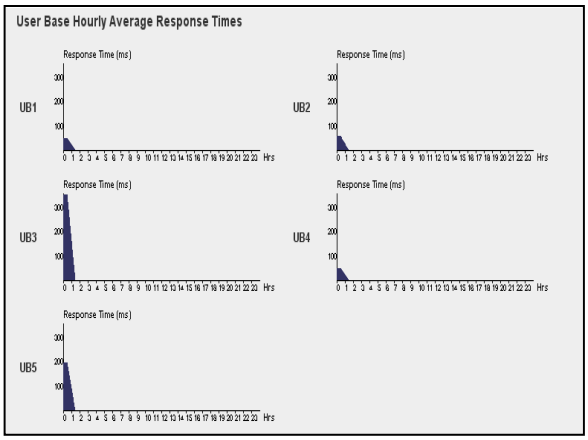


Fig. 2 : A Graph of User Base Hourly Average Response time using RR scheduler

From the above graph it is analysed that there is a high response time leading to very high proceesing time and ultimately resulting in increased rejection in the number of the jobs submitted. Figure. 3. illustrates the number of jobs rejected at the peak hours graphically. This figure further infers that with increrase in the submission rate of jobs, rejection rate further increases. As a supporting to the aforementioned TABLE V. gives the percentage of job rejected using RR scheduler.

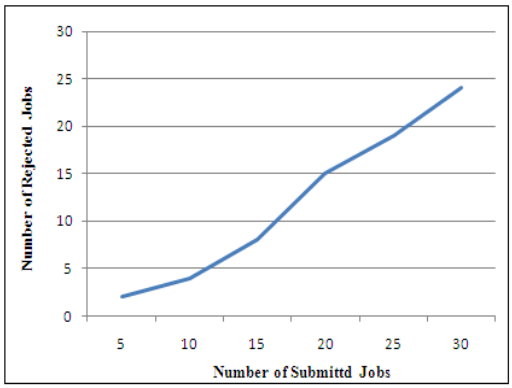


Fig. 3 : Graph Showing the number of jobs rejected using RR scheduler

TABLE V : PERCENTAGE OF JOB REJECTIONS USING RR SCHEDULER

No. of Jobs Submitted	Percentage of Jobs Rejected
5	40
10	40
15	53
20	75
25	76
30	80

As discussed earlier, the Load balancing is also one of important factor that influences the job rejection. The existing load balancing technique in cloud is Shortest Job First Scheduling (SJFS). This is further incorporated in the Load Balancer to analyze its efficiency in reducing the job rejections, in which the job with less burst time will be scheduled first by the load balancer.

TABLE VI. depicts the list of jobs received with their IDs and the burst time. According to the SJFS technique, the job with less burst time first gets scheduled for allocation. Therefore the drawback of this algorithm is that the jobs with higher burst time will be starved and gets rejected if the job is starving for a longer time and ultimately degrading the system performance.

TABLE VI : A SAMPLE OF JOBS RECEIVED

Job ID	Arrival time (hrs)	Burst time (hrs)
1	0	8
2	1	4
3	3	6
4	5	2
5	6	5

The scheduling order for the jobs shown in TABLE VI. will be Job 1, Job 4, Job 2, Job 5, Job 3 and the TALBE VII. depicts the Response time obtained from using the SJFS technique for Load Balancing and Figure 4. shows this graphically.

TABLE VII : RESPONSE TIME OBTAINED USING SJFS FOR LOAD BALANCING

Job ID	Response Time (hrs)
1	0
4	3
5	8
2	9
3	16

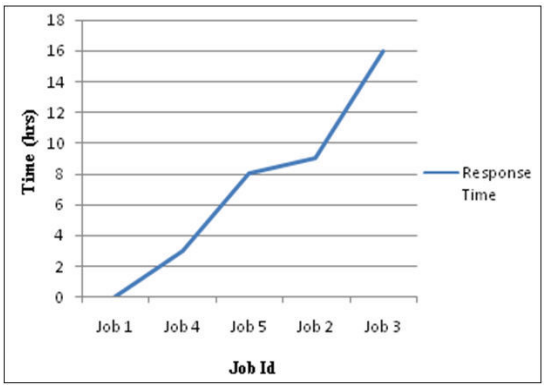


Fig. 4 : A Graph showing the Response Time obtained using SJFS technique for Load Balancing

From TABLE VII. and Figure. 4. it is analyzed that the Job 3 has got very high response time and it will be in starvation for a very long time resulting in rejection. Hence, with this strategy, increase in the number of job arriving to the system increases the response time and thereby results in starvation and job rejections. Figure 5. illustrates the number of jobs rejected at peak hours graphically by using the SJFS technique for load balancing. TABLE VIII. shows the percentage of job rejections using SJFS technique for load balancing. It is indicated from this table that with increased submission of jobs increases job rejections resulting in business loss to the cloud service provider.

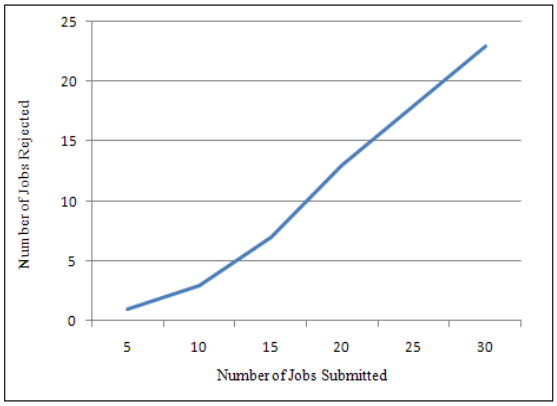


Fig. 5 : A Graph showing the number of jobs rejected using the SJFS technique for Load Balancing

TABLE VIII : PERCENTAGE OF JOBS REJECTED USING SJFS TECHNIQUE FOR LOAD BALANCING

No. of Jobs Submitted	Percentage of Jobs Rejected
5	20
10	30
15	47
20	65
25	72
30	77

Another important challenge to reduce job rejections in the cloud environment is to avoid deadlocks among the jobs which are contending for the same resource at the same instance of time [15]. This issue can be handled through load balancing by promoting VM migrations such that jobs will be migrated from the over loaded VM to the under-utilized VM by comparing the wait time and the hop time. Henceforth, the availability of resources increases and the deadlock is avoided. Currently, in the cloud computing domain there is no such load balancing approach that can avoid deadlocks, which reduces job rejection.

V. CONCLUSION

Cloud Computing is a promising technology to support IT organizations in developing cost, time and resource efective products. Since, Cloud computing is a pay-go-model, it is necessary to reduce job rejections at the peak hors inoredr to improve the business performance of the cloud system. The job rejection can be reduced through job scheduling and load balancing tecjnique.

This investigation, analyzes the efficiency of Round Robin (RR) scheduling strategy used for job scheduling and Shortest Job First Scheduling (SJFS) technique used for load balancing to reduce the job rejections. It is evident from this analysis that both the starategies can be further enhanced inoder to reduce job rejections, which is necessary to accelerate the working and business performance of the clud system. Further, yo avoid deadlocks in the cloud system it is recommended to have an effective load blancing approach that can efficiently allocate the resources to resolve the dealdlocks.

REFERENCES

[1] Selvarani, S., Sadhasivam, G.S.: Improved Cost-Based Algorithm for Task Scheduling in Cloud Computing. In: 2010 IEEE, pp.1-5 (2010)

[2] Mousumi Paul, Goutam Sanyal: Task-Scheduling in Cloud Computing using Credit Based Assignment Problem. In: International Journal on Computer Science and Engineering (IJCSE), Vol. 3, Issue. 10, October 2011 pp. 3426- 3430 (2011)

[3] Xiaoqiao Meng, Canturk Isci, Jeffrey Kephart, Li Zhang, Eric Bouillet, Dimitrios Pendarakis: "Efficient Resource Provisioning in Compute Clouds via VM Multiplexing". In ICAC'10, June 7–11, 2010, Washington, DC, USA. Copyright 2010 ACM 978-1-4503-0074-2/10/06

[4] Hao Liu, Shijun Liu, Xiangxu Meng, Chengwei Yang, Yong Zhang, "LBVS: A Load Balancing Strategy for Virtual Storage," 2010 International Conference on Service Sciences, pp. 257-262.

[5] Shu-Ching Wang, Kuo-Qin Yan, Wen-Pin Liao, Shun-Sheng Wang, "Towards a Load Balancing in a Three-level Cloud Computing Network," 2010 IEEE, pp. 108-113.

[6] Martin Randles, David Lamb, A. Taleb-Bendiab, "A Comparative Study into Distributed Load Balancing Algorithms for Cloud Computing", 2010 IEEE 24th International Conference on Advanced Information Networking and Applications Workshops, pp. 551-556.

[7] Zenon Chaczko, Venkatesh Mahadevan, Shahrzad Aslanzadeh, Christopher Mcdermid, "Availability and Load Balancing in Cloud Computing," 2011 International Conference on Computer and Software Modeling IPCSIT vol.14 (2011), pp. 134-140, © (2011) IACSIT Press, Singapore.

[8] Nair, T.R.Gopalakrishnan, Vaidehi, M., "Efficient resource arbitration and allocation strategies in cloud computing through virtualization," 2011 IEEE International Conference on Cloud Computing and Intelligence

Systems September 2011, Beijing, China, pp. 397-401.

[9] Jiahui Jin, Junzhou Luo, Aibo Song, Fang Dong, and Ranqun Xiong, "BAR: An Efficient Data Locality Driven Task Scheduling Algorithm for Cloud Computing", 2011 11th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, pp. 295- 304.

[10] Yagiz Onat Yazir, Chris Mathews Roozbeh Farahbod "Dynamic Resource Allocation in Computing Clouds using Distributed Multiple Criteria Decision Analysis", IEEE Third International Conference on Cloud Computing, pp. 91-98, 2010.

[11] Daniel Warneke and Odej Kao "Exploiting Dynamic Resource Allocation for Efficient Parallel Data Processing in the Cloud". In IEEE Transactions on Parallel and Distributed Systems, pp 1-14, 2011.

[12] Jiyin Li, Meikang Qiu, Jain-Wei Niu, YuChen, Zhong Ming "Adaptive Resource Allocation for Preempt able Jobs in Cloud Systems".IEEE International Conference on Intelligent Systems Design and Applications, pp. 31-36, 2010.

[13] Rajkumar Buyya and Karthik Sukumar "Platforms for Building and Deploying Applications for Cloud Computing" CSI Communication, pp. 6-11, 2011

[14] T. R. Gopalakrishnan Nair, M. Vaidehi, K.S. Rashmi, V. Suma, "An Enhanced Scheduling Strategy to Accelerate the Business performance of the Cloud System, " Porc. InConINDIA 2012, AISC 132, pp. 461-468, © Springer-Verlag Berlin Heidelberg 2012.

[15] T. R. Gopalakrishnan Nair, Vaidehi. M, Suma. V, "Improved Strategies for Enhanced Business Performance in Cloud-based IT Industries", 26th Indian Institute of Engineers Congress, IEI Congress-2011, Bangalore, December,2011

◈◈◈

Intelligent Energy Management System for Residential Buildings Based on ZigBee Technology

Sangeeta Modi, Ritu Naiya & Shaik Shabana

Electrical and Electronics Department, PES Institute of Technology, Bangalore, India
E-mail: smodi@pes.edu

Abstract - Energy consumption in residential buildings account for 20 to 40 per cent of total energy consumed in a country and therefore represents a significant and potential source of energy savings. An Intelligent Energy Management System can contribute to major reductions of energy use in hundreds of millions of buildings. This paper gives an overview of sensor technology and wireless networks in the development of an intelligent energy management system for residential buildings (IEMSRB). This technology has ample potential to change the way we live and work. In this paper ZigBee is used as a communication medium in building intelligent energy management system. From the prototype setup, it is shown that ZigBee is a suitable technology to be adopted as the communication infrastructure in energy management system for residential buildings .The performance analysis discussed in this paper verifies the effectiveness of using ZigBee in energy management system. The novelty of the present scheme is its ability to save the energy and improve the performance as it learns and gains more experience in real-time operations. Results also demonstrate that the proposed scheme can achieve the minimum electricity cost for residential customers. The proposed system can be installed and maintained in residential environments with ease.

Keywords - *Energy Management System; ZigBee; Sensor and Actuator Networks, Intelligent Energy Management System for Residential Buildings.*

I. INTRODUCTION

Energy was once a commodity which most enterprises did not have that much control of in the past. In this day and age, energy control is now a top priority. It is important for all of us to have the responsibility and make sure that we have a good energy system. The energy system includes energy of water, electricity, gas, air, and steam. The energy system cannot only affect our corporation or industry, but also affect the environment around our corporation or industry. An intelligent energy management system can contribute to major reductions of energy use in hundreds of millions of buildings. Energy savings and user happiness are two major design considerations for intelligent home system.Intelligent homes in a building must fulfill four basic requirements. First, they must facilitate a safe, convenient, and healthy lifestyle. Secondly, they must be environmentally sustainable. Thirdly, they must promote comfort. And finally, an intelligent home must provide an efficient workspace to its occupants. The most effective way to reduce energy is to turn devices off. The second most effective way is to turn them down. An automated control system can do both for consumer based on factors such as occupancy, available daylight and time of day. Removing the wires from the controls provides additional benefits, including greater flexibility in where controls can be placed, and significant savings in installation by avoiding the expense and disruption of wiring.This paper describes the development of an intelligent energy management system for residential buildings using the concept of a sensor technology and wireless network. In this paper ZigBee is used as a communication medium in energy management systems. By using this energy management system, it is possible to see and control the energy system of various devices. Wireless sensor technology is fast replacing wired technology in almost all the fields, because it is less costly and also because it is more efficient as compared to wired networks [1]. Sensors measure multiple physical properties and include electronic sensors, biosensors, and chemical sensors. These sensors can thus be regarded as "the interface between the physical world and the world of electrical devices, such as computers" [2].

In this paper ZigBee is used as a communication medium in energy management system which can be implemented in building, household, research laboratory and so on. The rest of the paper is organized as follows: Section II describes the sensor and actuator networks. In Section III, the ZigBee network is discussed. Section IV

discusses the proposed intelligent system for the residential building. Section V describes the implementation details and performance analysis, followed by a conclusion in Section VI

II. WIRELESS SENSOR AND ACTUATOR NETWORKS

The whole point of a wireless network is to send reliable data between nodes in the network. Wireless sensor and actuator networks (WSANs) are networks of nodes that sense and potentially also control their environment. They communicate the information through wireless links "enabling interaction between people or computers and the surrounding environment". The data gathered by the different nodes is sent to a sink which either uses the data locally, through for example actuators, or which "is connected to other networks (e.g. the Internet) through a gateway. Sensor nodes are the simplest devices in the network. A sensor node typically consists of five main parts: one or more sensors gather data from the environment. The central unit in the form of a microprocessor manages the tasks. A transceiver communicates with the environment and a memory is used to store temporary data or data generated during processing. Fig 1 shows architecture of a sensor node. To assure a sufficiently long network lifetime, energy efficiency in all parts of the network is crucial. Due to this need, data processing tasks are often spread over the network, *i.e.* nodes co-operate in transmitting data to the sinks. Fig 2 shows the most important fields of application. If compared the performance with wired Local Area Network (LAN), it is generally accepted that wired LAN network offers higher speed than wireless LAN network.

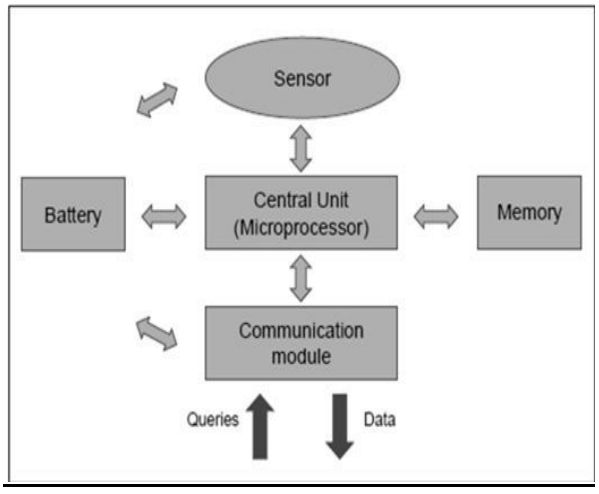


Fig 1 : Architecture of a sensor node

Fig 2 : Application of wireless sensor networks

The distinct in transmission speed is even more obvious when optical cable is being used in wired LAN network which transmission rate can easily reach up to 1Gbps or more. Although wireless communication system is less reliable, prone to interference and lower transmission rate at 54 Mbps, it is still being used by some power companies due to the advantages that offered only by wireless communication network [9,10]. LAN technologies connect different smart devices at customers' sites. These technologies can be classified into three main groups: wireless IEEE standards 802.x, wired Ethernet, as well as in-building power line communications Wireless IEEE standards include Wi-Fi (IEEE 802.11), WiMAX5 (IEEE 802.16), ZigBee (IEEE 802.15.4) and Bluetooth (IEEE 802.15.1). Based on [7], Table 1 shows a short description of strengths and weaknesses of these standards.

Table 1 : Comparison of IEEE Standards

IEEE Standard	Strengths	Weaknesses
ZigBee (IEEE 802.15.4)	• Low power requirements • Low implementation cost • Good scalability	• Limited range • Relatively low data rates • Possibly more secure than other standards
Wi-Fi (IEEE 802.11)	• Easy deployment and falling costs	• Only useful within the customer site • Additional security layers required
Bluetooth (IEEE 802.15.1)	• Permits higher data rates than ZigBee	• Limited maximum number of devices in a network • Security vulnerabilities

ZigBee is used to provide an efficient wireless communication standard for Home Area Networking

III. ZIGBEE NETWORKS

A. *Introduction and Characteristics*

Energy conservation, control, and safety are some of the prospects of ZigBee. Word ZigBee was originated from word Zigzag indicating cross-shaped network cables and Bee to indicate economical communication method. The name refers to the waggle dance of honey bees after their return to the beehive. The ZigBee network automatically figures out how to route the data from one node to another with the maximum chance of success. ZigBee networks have the following requirements and features: low power consumption, low cost, low packet throughput, lots of network nodes, low request on quality of service, security control, and high reliability [3]. ZigBee can be used in various applications such as HVAC controls, Lighting Controls, and Utility Networks. ZigBee consumes low electricity supply and can be configured to large scale sensor networks by integrating with sensor (Activity, light, temperature and humidity, etc) and transmitter/receiver devices. This type of structure is defined as foundation technology for sensing, monitoring and controlling. ZigBee has recognized as next generation short-distance wireless communication standard based on strong advantages including lowest costs, lowest energy consumption which can be last 2 years with 2 AA type batteries, scalability of up to 65,000 nodes, simple network configuration and reliability from immediate recovery function from data transmission errors. Especially, ZigBee supports multi-hop function to ensure highest transmission success rates.

B. *ZigBee Topologies*

ZigBee supports star, peer-to-peer i.e. mesh, and tree topologies [4]. In star topology, there are several nodes and a central coordinator [5]. Coordinator is the main part of star topology, as communication between nodes takes place through the coordinator. Nodes can communicate directly in peer-to-peer topology; Nodes can communicate directly in peer-to-peer topology, without the need of coordinator. In tree topology, network consists of a central coordinator node along with routers and other nodes [6].

IV. PROPOSED INTELLIGENT ENERGY MANAGEMENT SYSTEM FOR RESIDENTIAL BUILDINGS

Fig 3 shows the architecture of the proposed Intelligent Energy Management System for Residential Buildings (IEMSRB).Each home of the building has one living room, one bed room, one kitchen, one rest room and each section is equipped with necessary load as shown in Fig 3, one power outlet, and one ZigBee hub. The dimming light and the power outlet include a power measurement function to measure the power consumption. They report the information periodically to the ZigBee hub through ZigBee communication. Because home appliances are connected to the power outlet, their power usage can be acquired by the power measurement function of the power outlet. The ZigBee hub in the room gathers the power information reports of the light and the power outlet, and then it transfers the information to the home server. The home server analyzes the power information of all home appliances in each room. It displays the real-time active power consumption of each home appliance and the accumulated power consumption of each home appliance. A user can figure out which home appliance is unnecessarily turned on through the real-time active power consumption and how much power each home appliance consumes in this month through the accumulated power consumption. A user can also analyze the power usage of each room through the ZigBee hub. A user can access the home server and turn off unnecessarily turned on home appliances. The power outlet periodically monitors the power consumption of the connected home appliance. As soon as the monitored power consumption of the home appliance is below the threshold for the determined period, the power outlet automatically cuts off the AC power to reduce the standby power of a home appliance.

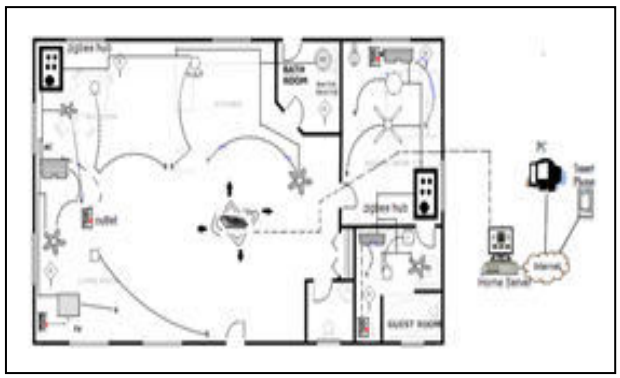


Fig 3 : Architecture of the proposed intelligent energy management

V. IMPLEMENTATION DETAILS AND PERFORMANCE ANALYSIS

To show the achievability of the proposed architecture an experimental based case study has been done on the system which is developed to demonstrate that smart, simple sensor devices can be used to manage, control and save energy in smart home in a smart building. We have developed a smart node that has sensing, processing and networking abilities. It is equipped with a microcontroller (8952) as shown in Fig 6 and a narrow-band radio frequency (RF) device that can support physical-layer functionalities of IEEE 802.15.4. Various optional sensor and actuator modules

can be equipped with this smart node with the help of a connector and directly controlled by the microcontroller in our smart node. Temperature sensor is included in the smart node. An experimental based case study has been done on the system which is developed to demonstrate that smart, simple sensor devices can be used to monitor activities of daily living and life style of person living in smart home in a smart building. The system has been tested by connecting the soldering iron to the temperature sensor as shown in Fig 4 and Fig 5.

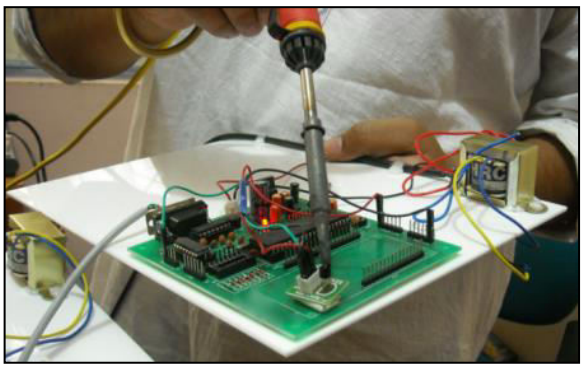


Fig 4 : System Testing with Soldering Iron

Fig. 5 : Temperature Measuring Unit

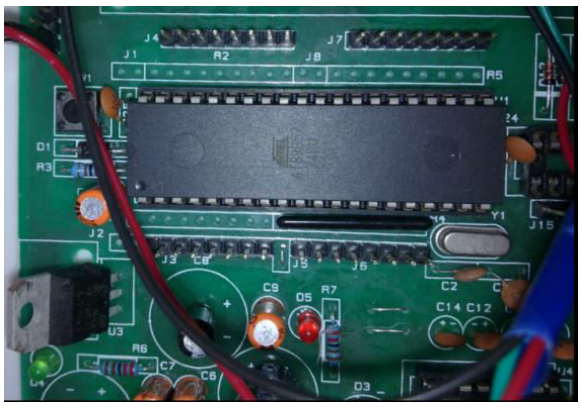


Fig 6 : Microcontroller with voltage regulator

The following observations were made

Table 2 : Temperature Sensor Data

Sr. No.	Temperature monitored (°C)	Buzzer sounded
1.	16	No
2.	20	No
3.	23	No
4.	28	No
5.	31	No
6.	33	No
7.	40	Yes
8.	120	Yes
9.	35	No
10.	55	Yes

Each time the buzzer was sounded the power supply to the heating element was cut off thereby saving on energy. This has a wide application in areas where water or other fluids need to be heated or maintained at a constant temperature. In the second conduct of experiment to show the saving in energy consumption using ZigBee based temperature sensor, we have included an air conditioner (AC) and two tube lights in the system. ZigBee based temperature sensor network senses the ambient temperature of the room. The desired critical temperature required in the room is 27°C and the air conditioner should not exceed desired level, however lower temperatures can be allowed. Wattage of a 1 ton air conditioner (AC) is 3500 watts. Table 3 to Table 10 show the data collected during various seasons and Electricity bill on monthly basis for the said duration

On a winter morning if the average ambient temperature in the morning (6:00 am to 6:00 pm) is 25°C and the average ambient temperature at the night is (6:00 pm to 6:00 am) is 22°C then there is no necessity of the usage of air conditioner both in the day and night, then the sensor network sounds an alarm.

Table 3 : Winter Data

Time Duration	No of Air Conditioners working	No of lights on	kW Consumption of loads in an hour	kWh consumption for the duration specified
6:00 am – 6:00 pm	0	2	0.08	0.96
6:00 pm – 6:00 am	0	2	0.08	0.96
Daily energy consumption				1.92

Monthly energy consumption = 1.92*30 = 57.6 kWh

Table 4 : Electricity Bill during winter

Description	Units Consumed	Rate (Rs / unit)	Total Charges (Rs)
Tariff Charges	57.6	2.90	167.04
Electricity Duty	57.6	0.09	5.184
Total Monthly Bill			172.224

On an autumn morning, if the average ambient temperature in the morning (6:00 am to 6:00 pm) is 29°C and the average ambient temperature at the night is (6:00 pm to 6:00 am) is 27°C then there is no necessity of the usage of air conditioner in the night, however AC is required in the morning duration, then the sensor network sounds an alarm

Table 5 : Autumn Data

Time Duration	No of Air Conditioners working	No of lights on	kW Consumption of loads in an hour	kWh consumption for the duration specified
6:00 am – 6:00 pm	1	2	3.58	42.96
6:00 pm – 6:00 am	0	2	0.08	0.96
Daily energy consumption				43.92

Monthly energy consumption = 43.92*30 =1317.6 kWh

Table 6 : Electricity bill during autumn

Description	Units Consumed	Rate (Rs / unit)	Total Charges (Rs)
Tariff Charges	1317.6	2.90	3821.04
Electricity Duty	1317.6	0.09	118.584
Total Monthly Bill			3939.624

On a summer morning, if the average ambient temperature in the morning (6:00 am to 6:00 pm) is 32°C and the average ambient temperature at the night is (6:00 pm to 6:00 am) is 28°C then if only AC is sufficient to maintain the desired level at both day and night time then the energy consumption is as follows:

Table 7 : Summer Data

Time Duration	No of Air Conditioners working	No of lights on	kW Consumption of loads in an hour	kWh consumption for the duration specified
6:00 am – 6:00 pm	1	2	3.58	42.96
6:00 pm – 6:00 am	1	2	3.58	42.96
Daily energy consumption				85.92

Monthly energy consumption = 85 .92*30=2577.6 kWh

Table 8 : Electricity Bill during summer

Description	Units Consumed	Rate (Rs / unit)	Total Charges (Rs)
Tariff Charges	2577.6	2.90	7475.04
Electricity Duty	2577.6	0.09	231.98
Total Monthly Bill			7707.024

In the absence of ZigBee based temperature control: If the average ambient temperature in the morning (6:00 am to 6:00 pm) is 32°C and the average ambient temperature at the night is (6:00 pm to 6:00 am) is 28°C and the ZigBee based sensor network will not control the temperature and both the ac's and light loads will be in operation. However it is not necessary that the temperature is always more than the desired level. This case is taken to explain the total kWh consumption when there is no control of temperature.

Table 9 : Data in the absence of ZigBee based sensor

Time Duration	No of Air Conditioners working	No of lights on	kW Consumption of loads in an hour	kWh consumption for the duration specified
6:00 am – 6:00 pm	2	2	7.08	84.96
6:00 pm – 6:00 am	2	2	7.08	84.96
Daily energy consumption				169.92

Monthly energy consumption=169.92*30 = 5097.6 kWh

Table 10 : Electricity bill in the absence of ZigBee based sensor

Description	Units Consumed	Rate (Rs / unit)	Total Charges (Rs)
Tariff Charges	5097.6	2.90	14783.04
Electricity Duty	5097.6	0.09	458.784
Total Monthly Bill			15241.824

Fig 7 gives graphical representation of the energy consumption in a room with and without the ZigBee based sensor in all the 3 seasons in a year. If February, March, April and May are assumed to be summer seasons June, July, August, September are autumn and October, November, December, and January are assumed to be winter seasons in tropical regions then the graphical description to explain energy saving is shown below. Graphical data that is plotted is tabulated below in Table 11.

Table 11 : Load Data on Monthly Basis

Months	Seasons	Monthly kWh consumption using ZigBee based temperature sensor	Monthly kWh consumption in absence of ZigBee based temperature sensor
January	Winter	57.6	5097.6
February	Summer	2577.6	5097.6
March	Summer	2577.6	5097.6
April	Summer	2577.6	5097.6
May	Summer	2577.6	5097.6
June	Autumn	1317.6	5097.6
July	Autumn	1317.6	5097.6
August	Autumn	1317.6	5097.6
September	Autumn	57.6	5097.6
October	Winter	57.6	5097.6
November	Winter	57.6	5097.6
December	Winter	57.6	5097.6

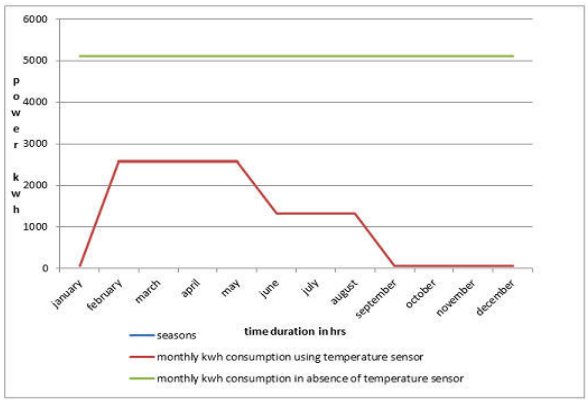


Fig. 7 : Monthly kWh consumption with and without ZigBee based sensor

VI. CONCLUSION

The system has been tested by connecting the ZigBee based temperature sensor to the following appliances - soldering iron, air conditioner and tube lights. Table 3 to Table 10 show the data collected during various seasons and Electricity bill on monthly basis for the said duration .Fig 7 shows the monthly kWh consumption of energy with and without ZigBee based sensors. This captured data can help us to identify load pattern and energy saving in smart homes of an intelligent building. The proposed system can be installed and maintained in residential environments with ease.A ZigBee based Intelligent Energy Management System for residential building system can provide significant cost savings in a building environment, great level of flexibility and control for the building administrators and great comfort for the occupants.

REFERENCES

[1] Baronti P, Pillai P, Chook V, Chessa S, Gotta A, Hu Y. Wireless sensor networks: A survey on the state of the art and the 802.15.4 and ZigBee standards. Computer Communications 2007; 30: 1655–1695.

[2] Wilson, J. (2008), Sensor Technology Handbook, Newnes/Elsevier, Oxford.

[3] Gorday P, Hester L, Callaway E. Home Networking with IEEE 802.15.4: A Developing Standard for Low-Rate Wireless Personal Area Networks. IEEE Communications Magazine 2002; 70-78.

[4] Getting Started with ZigBee and IEEE 802.15.4. Daintree Networks 2008.

[5] ZigBee\ZigBee.htm: Wikipedia

[6] P. Kinney, ZigBee Technology: Wireless Control that Simply Works, White Paper dated 2 October 2003.

[7] Electric Power Research Institute (EPRI, 2006), IntelliGrid[SM] Consumer Portal Telecommunications Assessment and Specification, Technical Report, Palo Alto, CA.

[8] ZigBee: The Choice for Energy Management and Efficiency – ZigBee White Paper - June 2007

[9] The magazine of the continental automated buildings association spring 2008 volume 5, number 1

[10] X. D. Zhang, Y. Gao, G. Y. Zhang, and G. Bi, "CDMA2000 cellular network based SCADA system," in Proc. IEEE International Conference on Power System Technology (PowerCon 2002), vol. 2, pp. 1301-1306, 2002.

Privacy-Preserving Data Mining

G. Rekha & K.Tejaswini

E-mail : rekha@kmit.in & tejaswini.kurakula@gmail.com

Abstract **-** The field of privacy has seen rapid advances in recent years because of the increase in the ability to store data. In particular, recent advances in the data mining field have lead to increased concerns about privacy. While the topic of privacy has been traditionally studied in the context of cryptology and information security, recent emphasis on data mining has lead to renewed interest in the field. A number of algorithmic techniques have been designed for privacy-preserving data mining. In this paper, we provide a review of state-of-art methods for privacy preserving. We also discuss the computational and theoretical limits associated with privacy- preservation over high dimensional data sets.

I. INTRODUCTION

The problem of privacy-preserving data mining has become more important in recent years because of the increasing ability to store personal data about users and the increasing sophistication of data mining algorithms to leverage this information. A number of techniques such as randomization and *k*-anonymity have been suggested in recent years in order to perform privacy-preserving data mining. The problem has been discussed in multiple communities such as database community, the statistical disclosure control community and the cryptography community. In some cases, the different communities have explored parallel lines of work which are quite similar.

The key directions in the field of privacy-preserving data mining are as follows:

1.1 Privacy-Preserving Data Publishing. This techniques lend to study different transformation methods associated with privacy. These techniques include methods like randomization, *k*-anonymity and *l*-diversity. Another related issue is how the perturbed data can be used in conjunction with classical data mining methods such as association rule mining. Other related problem of studying the different definitions of privacy, and how to compare them in terms of effectiveness in different scenarios.

1.2 Changing the results of Data Mining Applications to preserve privacy. In many cases, the results of data mining applications such as association rule or classification rule mining can compromise the privacy of the data. This has spawned the privacy in which the results of data mining algorithms such as association rule mining are modified in order to preserve the privacy. A classic example of such techniques are association rule hiding methods, in which some of the rules are suppressed.

1.3 Query Auditing. Such methods are related to previous case of modifying the results of data mining algorithms. Either modifying or restricting the results of queries .

1.4 Cryptographic Methods for Distributed Privacy. In many cases , the data may be distributed across multiple sites, and the owners of the data across these sites may wish to compute a common function. A variety of cryptographic protocols may be used in order to communicate among the different sites, so that secure function computation is possible without revealing sensitive information. While the individual entities may not desire to share their entire data sets, they may consent to limited information sharing with use of a variety of protocols. The overall effect of such methods is to maintain privacy for individual entity.

1.5 Theoritical Challenges in high Dimensionality: Real data sets are usually extremely high dimensional, and this makes the process of privacy preservation extremely difficult both from a computational and effectiveness point of view showing optimal *k*-anonymitization is NP hard.

1.6 Downgrading Application Effectiveness: In many cases, even though the data may not be available, the output of applications such as association rule mining, classification or query processing may result in violations of privacy. This has lead to research in downgrading the effectiveness of applications by either data or application modifications.

2. Privacy-preserving data Mining Algorithms

In this section, we will discuss the key stream problems and will discuss the challenges associated with each problem. We will also discuss an overview of the material covered as mentioned below:

2.1 General Survey: An overview of the different techniques and how they relate to one another. The individual topics will be covered in sufficient detail. More detailed discussions are deferred to future chapters which contains descriptions of different data mining algorithms.

2.2 Statistical Methods for Disclosure Control: This method includes *k*-anonymity, swapping, randomization, micro-aggregation and synthetic data generation. The idea is to give the reader an overview of the common themes in privacy preserving data mining by different communities.

2.3 Measures of Anonymity: There are a very large number of definitions of anonymity in the privacy-preserving data mining field. This is partially because of the varying goals of different privacy-preserving data mining algorithms. For example, methods like *k*-anonymity ,*l*-diversity and *t*-closeness are all designed to prevent identification, though the final goal is to preserve the underlying sensitive information. This thus provides an overview and perspective of the different ways in which privacy could be defined and its advantages.

2.4 The *k*-Anonymity Method. The motivating factor behind this technique is that many attributes in the data can be often pseudo-identifiers which can be used in conjunction with public records in order to uniquely identify them. This is because combinations of record attributes can be used to exactly identify individual records The idea of *k*-anonymity is to reduce the granularity of representation of the data in such a way that a given record be distinguished from at least (k-1) other records with the help techniques such as generalization and suppression. This method of generalisation the attribute values are generalized to a range in order to reduce the granularity of representation and to maintain the data precision as possible for anonymity.

The k-optimize algorithm assumes an ordering some of the quasi identifier where the values of the attributes are discretized into intervals. Each such grouping is called item. For a given attribute, the corresponding items are also ordered. An index is created using these pairs. This set enumeration tree is a systematic enumeration of all possible generalizations. The root of the node is null and every successive level of the tree is constructed by appending one item which is lexicographically larger than all the item set of that node. A node of a tree can be pruned when it is determined that no descendant of it could be optimal. This can be done by computing a bound on the quality of all descendants of that node and comparing it to the quality of the current best solution obtained during the during the traversal process. A branch and bound technique can be used to successively improve the quality of the solution during the process. Eventually it is possible to terminate the process at maximum computational time.

Incognito method has been proposed for computing a k-minimal generalizations with the use of bottom up breadth first search aggregation along domain. It again computes generalization pairs, again pruning those pairs which do not satisfy the k-anonymity constraints. In general Incognito algorithm computes (i+1) dimensional generalizations candidates from the i-dimensional generations and remove all those which donot satisfy.

Clustering is the technique that generates pseudo-data that uses principle data analysis of the behaviour of the records within a group. It has been the most effective approach for the problem of classification and the pseudo data layer provides additional security

2.5 The Randomization Method. This technique is used for privacy preserving in which noise is added to the data in order to mask the attribute values of record that is sufficiently large so that the individual records cannot be recovered. This technique uses data distortion method in order to create private representations of the records. The individual records cannot be covered, but only aggregate distributions can be recovered. These aggregate distributions can be used for data mining purposes. Two kinds of perturbations are possible with the randomization method.

The method of randomization can be described as follows. Consider a set of data records denoted by

$X = \{ x_{1,\ldots\ldots,}x_2\}$ for x_n X, we add a noise component which is drawn from the probabilistic distribution $f_y(Y)$. These noise components are drawn independently and are denoted $y_1.....y_n$. thus, the new set of distorted records are denoted by $x_1+y1_{\ldots\ldots}x_n+y_n$. We denote the new set as $z_{1\ldots\ldots\ldots\ldots}z_n$ in general, it is assumed that the variance of the added noise is large enough, so that the original record values cannot be easily guessed from the distorted data.

One key advantage of this method is that it is relatively simple, and does not require knowledge of the distribution of other records in the data. Therefore, this method can be implemented at data collection time and does not require the use of trusted server containing all the original records in order to perform the anonymization process and other OLAP applications.

2.5.1 Additive Perturbations: randomised noise is added to the records. The overall data distributions can be recovered from the randomised records. Data mining and management algorithms redesigned to work with these data distributions.

2.5.2 Multiplicative Perturbations: The random projection or random rotation techniques are used in order to perturb the records. It is a good effort for privacy preserving data mining and they show how to use multi-dimensional projections in order to reduce the dimensionality of data. This technique preserves the inter-record distances and the transformed records can be used in conjunction with other applications.

As in case of additive perturbations, multiplicative perturbations are not entirely safe from adversarial attacks. In general, if the intruder has no prior knowledge of the data, then it is relatively difficult to attack the privacy of transformation. There are two kinds of attacks:

2.5.2.1 Known Input- Output Attack: The attacker knows some linearly independent collection of records. In such case linear algebra techniques can be used to reverse-engineer the nature of privacy preserving transformations

2.5.2.2 Known Sample Attack: The attacker has a collection of independent data samples from the distribution from which the original data was drawn. Principal component analysis can be used to reconstruct the behaviour of the original data.

3. In addition, privacy-preservation deals with the issue of adversarial attacks and vulnerabilities of these methods.

3.1 Quantification of Privacy. A key issue in measuring the security of different privacy-presentation methods is the way in which the underlying privacy is quantified. The idea in privacy quantification is the measure the risk of disclosure for a given level of perturbation.

The quantity used to measure privacy should indicate how closely the original value of an attribute can be estimated. If the original value can be estimated with c% confidence to lie in the interval [$1, $2] then the value with ($1, $2) defines the amount of privacy at c% level. For suppose the perturbing attribute value is uniformly distributed in interval of width 2$, then $ is the confidence level. However, this simple method of determining privacy can be subtly incomplete in some situations.

3.2 Utility Based Privacy-Preserving Data Mining. Most privacy-preserving data mining apply a transformation which reduces the effectiveness of the underlying data when it is applied to data mining algorithms. In fact ,there is a natural trade off between privacy and accuracy. The issue of designing utility based algorithms to work effectively with certain kinds of data mining problems.

3.3 Mining Association Rules Under Privacy Constraints. Since association rule mining is one of the important problems in data mining. There are two aspects to the privacy-preserving association rule mining problem:

3.3.1 When the input to the data is perturbed , it is a challenging problem to accurately determine the association rules on the perturbed data.

3.3.2 A different issue is that of output association rule privacy. In this case we try to ensure that none of the association rules in the output result in leakage of the sensitive data. This problem is referred as *association rule hiding* by the data base community, and that of *contigency table privacy- preservation* by the statistical community.

3.4 Cryptographic Methods for Information Sharing and Privacy. In many cases, multiple parties may wish to share aggregate private data, without leaking any sensitive information at their end. For example , different super stores with sensitive sales data may wish to coordinate themselves in knowing aggregate trends without leaking their individual stores. This requires *secure and cryptographic protocols* for sharing the information across different parties. The data may be distributed in two ways across different sites:

3.4.1 Horizontal partitioning: The different sites may require different set of records containing the same attributes.

3.4.2 Vertical portioning: The different sites may require different set of records containing the same set of records.

4.1 Privacy Attacks. It is useful to examine the different ways in which one can make adversarial attacks on privacy-transformed data. This helps in designing more effective privacy-transformation methods. Some methods which can be used in order to attack the privacy of the underlying data include SVD-based methods, spectral filtering methods and the background knowledge attacks. The broad idea in this technique is the correlation structure in the original data can be estimated fairly accurately even after noise addition. Once the relation is determined, one can even try to remove the noise in the data in such a way that it fits the aggregate correlation structure of the data. It has been shown that such techniques can reduce the privacy of the perturbation process significantly since the noise removal results in values that are close to original values.

The second kind of attack is with the use of public information. Consider a record $X=(x_1.....x_n)$, which is perturbed to $Z=(z1.....z_n)$. consider the public record $W=(w_1.....w_n)$. then the potential perturbation of Z with respect to W is given by $(Z-W)=(z_1-w_1.....z_d-w_{d)}$. Each of these values should fit the distribution f(Y). If it is known that the public data set always includes X, then the maximum likelihood fit can be provide a high degree of certainty in identifying the correct record, especially in cases where n is large.

4.2 Query Auditing and Inference Control. Many private databases are open to querying. This can compromise the security of the results, when the adversary can use different kinds of queries in order to undermine the security of the data. For example, a combination of range queries can be used in order to narrow down the possibilities for that record. Therefore, the results over multiple queries can be combined in order to uniquely identify a record, or at least the uncertainty in identifying it. There are two primary methods for preventing this kind of attack:

4.2.1 Query output Perturbation: we add noise to the output of the query result in order to preserve privacy.

4.2.2 Query Auditing: We choose to deny a subset of the queries so that the particular combination of queries cannot be used in order to violate the privacy.

4.3 Privacy and the Dimensionality Curse. In recent years, it has been observed that many privacy-preservation methods such as *k- anonymity* and randomization are not very effective in the high dimensional case.

4.4 Personalised Privacy Preservation. In many application, different subjects have different requirements for privacy. Here we construct anonymizations of the data such that different records have different levels of privacy. The methods involved in it are *condensation and conventional* approach .

4.5 Privacy-Preservation of Data Streams. A new topic in the area of privacy-preserving data mining is that of data streams, in which data grows rapidly at unlimited rate. In such cases, the problem of privacy-preservation is quite challenging since the data is being released incrementally. In addition, the fast nature of data streams obviates the possibility of using the past history.

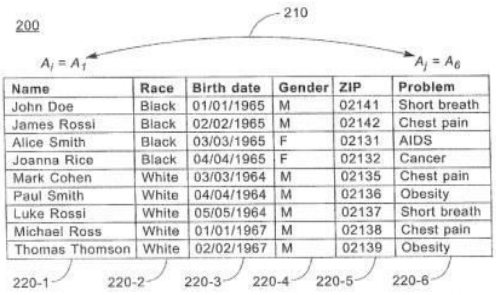
4.6 Utility Based Privacy Preservation. The broad idea is to ameliorate the curse of dimensionality by separately publishing marginal tables containing attributes which have utility but are also problematic for privacy preservation.

Figures and Diagrams

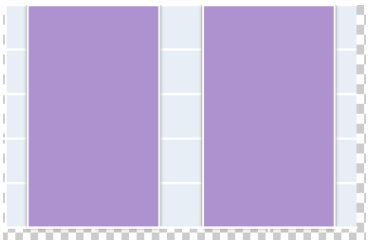
Diagram representing the data privacy

Table 1

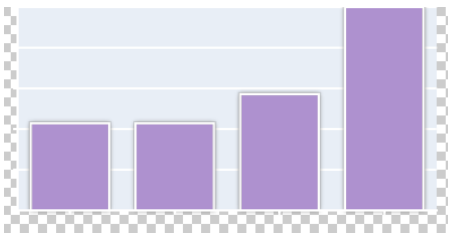
	Integrity	Confidentiality	Privacy
High	Untrusted	Secret	Private
Low	Trusted	Public	Public
Downgrading	Endorsement	Declassification	Anonymization



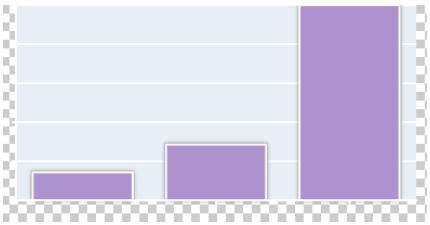
Patent applications-1



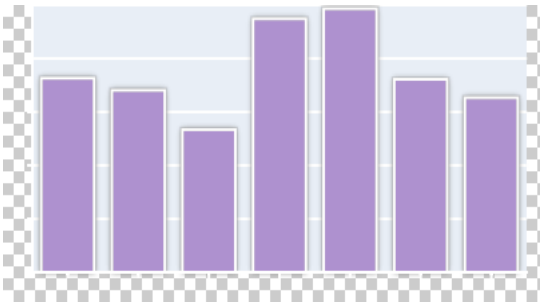
Patent applications-2



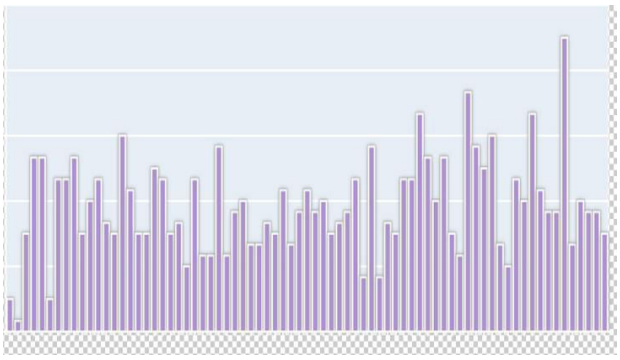
Patent applications-3



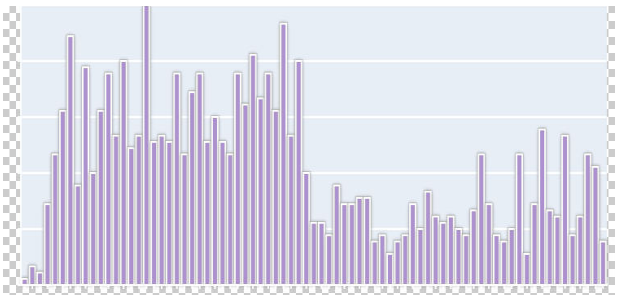
Patent applications-4



Patent application classes-Prevention of unauthorized use of data including prevention of piracy, piracy violations, or unauthorised data modifications



Patent application in all subclasses- Prevention of unauthorized use of data including prevention of piracy, piracy violations, or unauthorised data modifications



CONCLUSION AND SUMMARY

The broad areas of privacy are as follows:

1. Privacy-Preservation Data Publishing: This corresponds to sanitizing the data, so that its privacy remains preserved.

2. Privacy-Preserving applications: this corresponds to designing data management and mining algorithms in such a way that the privacy remains preserved. Some examples include association rule mining, classification and query processing.

3. Utility Issues: Since the perturbed data may often be used for mining and management purposes, its utility needs to be preserved. Therefore, the data mining and privacy transformation techniques need to be designed effectively, so to preserve the utility of the results.

4. distributed Privacy, Cryptography and Adversarial collaboration: This corresponds to secure communication protocols between trusted parties, so that information can be shared effectively without revealing sensitive information about particular parties.

REFERENCES

[1] Agrawal R. , Srikanth R. Privacy-Preserving data Mining. ACM SIGMOD conference, 2000.

[2] Aggarwal C. C., Yu P.S.;A Condensation approach to Privacy Preserving data mining. EBDT conference,2004.

[3] Bayrado R J.,Agrawal R.data Privacy through optimal *k*-anonymization.

[4] Blum A., Dwork C., mcSherry F.,Nissim K.: Practical privacy: The SuLQ framework.

[5] Aggarwal C. C.,Yu P.S.: On variable Constraints in privacy-preservating data mining

[6] Agrawal D.Aggarwal C.C On the design and quantification of privacy preserving data mining algorithms.

[7] Aggarwal C.C. On randomization, Public Information, And the curse of Dimensionality.

[8] Samarati P.,Sweeney L. Protecting privacy when disclosing Information:

[9] Verykios V. S.,Elmagarmid A., Bertino E., Saygin Y.,Association Rule Hiding.IEEE transaction on Knowledge and data engineering.

[10] Meyerson A., Williams R. On the complexity of *k*-anonymity.

[11] Rizvi S., Haritsa J.Maintaining data privacy in association Rule mining.

◈◈◈

FPGA Implementation of Red Algorithm For High Speed Pupil Isolation

[1]Prashanth M Ullagaddi, [2]K N Pushpalatha & [3]Aravind Kumar Gautam

[1&2] Dept of ECE, Dayananda Sagar College Of Engineering, Bangalore-78
[3]SD college Of Engineering, Muzaffar Nagar, U.P-54
E-mail : pullagaddi@gmail.com, knpdrs@gmail.com & drakgautam@gmail.com

Abstract - Iris recognition is an automated method of biometric identification that uses mathematical pattern-recognition techniques on video images of the irises of an individual's eyes, whose complex random patterns are unique and can be seen from some distance. Modern iris recognition algorithms can be computationally intensive, yet are designed for traditional sequential processing elements, such as a personal computer. However, a parallel processing alternative using Field Programmable Gate Array offers an opportunity to speed up iris recognition. Within the means of this project, iris template generation with directional filtering, which is a computationally expensive, yet parallel portion of a modern iris recognition algorithm, is parallelized on an FPGA system. An algorithm that is both accurate and fast in a hardware design that is small and transportable are crucial to the implementation of this tool. As part of an ongoing effort to meet these criteria, this method improves a iris recognition algorithm, namely pupil isolation. A significant speed-up of pupil isolation by implementing this portion of the algorithm on a Field Programmable Gate Array.

I. INTRODUCTION

Daugman created a very strong algorithm for iris detection based on Gabor wavelets .An alternate design, Ridge Energy Direction is based on spatial domain directional filters. This application has great military interest, and a small system on a chip design using a Field Programmable Gate Array (FPGA) would be ideal for carrying into the field or storing in a backpack [2]. FPGA devices also provide an attractive solution to computationally intensive applications because of their high density, high performance and complete configurability to support specific applications. An FPGA chip offers a combination of the flexibility of general purpose computers and hardware-based real-time processing of Application Specific Integrated Circuits (ASICs). An architecture design for FPGA technology can fully exploit the data and I/O parallelism in most image processing applications.

An FPGA chip offers a combination of the flexibility of general purpose computers and hardware-based real-time processing of ASICs. An architecture design for FPGA technology can fully exploit the data and I/O parallelism in most image processing applications. Our ultimate goal is to build each section of the RED algorithm using the most efficient method and combine them to create a single fast and accurate system for iris recognition.

The iris is a thin circular diaphragm, which lies between the cornea and the lens of the human eye. A front-on view of the iris is shown in Figure 1. The iris is perforated close to its center by a circular aperture known as the pupil. Formation of the iris begins during the third month of embryonic life. The unique pattern on the surface of the iris is formed during the first year of life, and pigmentation of the stroma takes place for the first few years. Formation of the unique patterns of the iris is random and not related to any genetic factors. Segmentation isolates the actual iris region in a digital eye image. The success of Segmentation depends on the imaging quality of eye images.

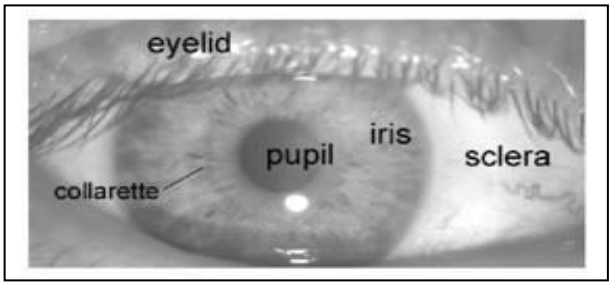


Figure 1. Front on view of iris

Segmentation normalizes the inconsistency caused by pupil dilation, varying imaging distance, rotation of the camera, head tilt, and rotation of the eye within the eye socket.

The most discriminating information present in an iris pattern is extracted by feature encoding. Only the significant features of the iris must be encoded so that comparisons between templates can be made. Matching the template that is generated in the feature encoding process will also need a corresponding matching metric, which gives a measure of similarity between two iris templates.

After determining the inner and outer boundaries and centre of the pupil, the iris is again transformed into polar Coordinates with the centre of the pupil as the point of reference, into a 120 row by 180 column image. In this process, the radial extent of the iris is normalized in order to account for pupil dilation. Each row in the unwrapped iris image represents an annular region surrounding the pupil, and the columns represent radial form information.

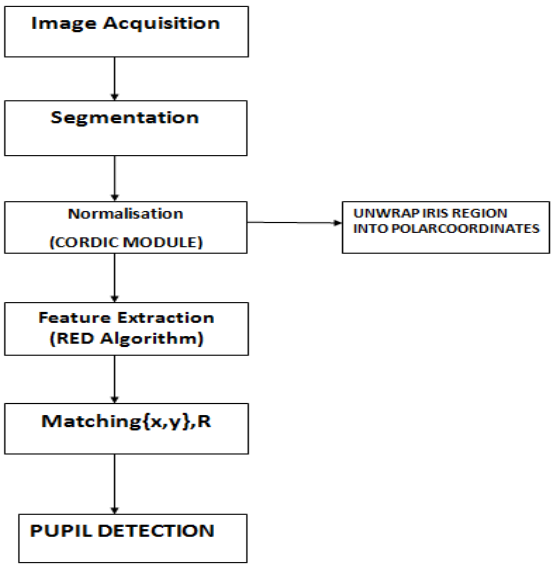


Figure 2: Processing flow diagram

II. RED ALGORITHM STEPS

Iris recognition requires four main steps:

i. Image capture;

ii. Pre-processing, which includes segmentation (isolating the iris from the image of the eye area), and usually a polar coordinate transform of the annular iris region into a rectangular image;

iii. Feature extraction, which generates an iris template; and

iv. Comparison of iris templates and a recognition (matching) decision. A number of methods of pre-processing and comparison are present.

A. Canny Edge Detection

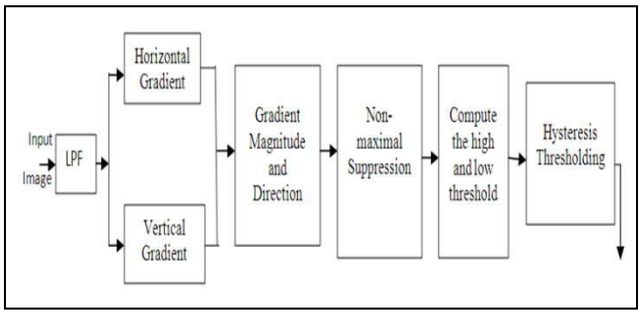


Figure 3. Block diagram of the canny edge detection algorithm

The Canny edge detector is predominantly used in many real-world applications due to its ability to extract significant edges with good detection and good localization performance. Unfortunately, the Canny edge detection algorithm contains extensive pre-processing and post-processing steps and is more computationally complex than other edge detection algorithms, such as Roberts, Prewitt and Sobel algorithms. Furthermore, it performs hysteresis thresholding which requires computing high and low thresholds based on the entire image statistics.[6] This places heavy requirements on memory and results in large latency hindering real-time implementation of the Canny edge detection algorithm.

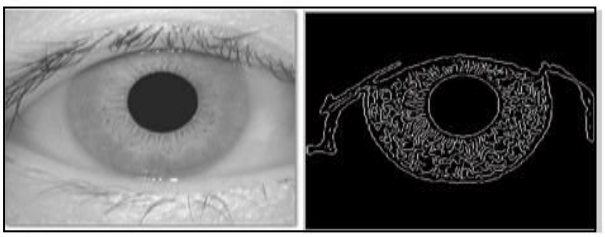


Figure 4. Photograph of the eye captured by a digital camera (a)and after Canny edge detection (b).

There are several steps required to achieve this end state. From the infrared image of the eye of a compliant participant (Fig. 4a), the iris must be extracted as accurately as possible. Several factors can cause interference such as the presence of the eyelid and eyelashes, reflections, and different angles of the eyeball. The application must be able to isolate and extract the pupil from within the iris. One of the means to do this in the RED algorithm is described as follows. The first step is to apply Canny edge detection (Fig. 4b). [1]Then the outline of the pupil is located using a search for circles of various radii. This paper speeds up the process of finding the inner iris border, using an FPGA to parallelize the search. This paper has been described a computational approach to edge detection. [4]The

success of the approach depends on the definition of a comprehensive set of goals for the computation of edge points. These goals must be precise enough to delimit the desired behaviour of the detector while making minimal assumptions about the form of the solution.

The main steps included are the one dimensional formulation. In the above steps three performance criteria are followed, Good detection, Good localization, Only one response, [8] Finding optimal detection by numerical optimization, A detector for step edge, [4]. An efficient approximation, two or more dimensions, need of multiple width, need of directional operators

Here the defined detection and localization criteria for a class of edges, and present mathematical forms for these criteria as functional on the operator impulse response. We have also found that the first two criteria are not "tight" enough, and that it is necessary to add a third criterion to circumvent the possibility of multiple responses to a single edge. Using numerical optimization, we derive optimal operators for ridge and roof edges. We will then specialize the criteria for step edges and give a parametric closed form for the solution.

In the process we will discover that there is an uncertainty principle relating detection and localization of noisy step edges, and that there is a direct tradeoff between the two. One consequence of this relationship is that there is a single unique "shape" of impulse response for an optimal edge detector, and that the tradeoff between detection and localization [9] can be varied by changing the spatial width of the detector. Several examples of the detector performance on real images will be given.

There was a direct tradeoff in detection performance versus localization, and this was determined by the spatial width. A detector was proposed which used adaptive thresholding with hysteresis to eliminate streaking of edge contours. The thresholds were set according to the amount of noise in the image, as determined by a noise estimation scheme. [9]This detector made use of several operator widths to cope with varying image signal-to-noise ratios, and operator outputs were combined using a method called feature synthesis, where the responses of the smaller operators were used to predict the large operator response.

The performance of the canny algorithm depends heavily on the adjustable parameters, σ, which is the standard deviation for the Gaussian filter, and the threshold values, 'T1' and 'T2'. σ also controls the size of the Gaussian filter. The bigger the value for σ, the larger the size of the Gaussian filter becomes. This implies more blurring, necessary for noisy images, as well as detecting larger edges. [4][6]As expected, however, the larger the scale of the Gaussian, the less accurate is the localization of the edge. Smaller values of σ imply a smaller Gaussian filter which limits the amount of blurring, maintaining finer edges in the image. Canny's edge detection algorithm is computationally more expensive compared to Sobel, Prewitt and Robert's operator. However, the Canny's edge detection algorithm performs better than all these operators under almost all scenarios.

B. Red Algorithm

The "energy" of the unwrapped iris image after contrast-limited adaptive histogram equalization is considered. Here, "energy" loosely refers to the prominence (pixel values) of the ridges that appear in the histogram equalized image, higher value reflects higher energy. This "energy" image is passed into each of two different 11 x 11 directional filters (a vertical filter and a horizontal filter). These filters are used to indicate the presence of strong ridges, and the orientation of these ridges.[3]

At every pixel location in the filtered image, the filter which provides the largest value of output is recorded and encoded with one bit to represent the identity of this directional filter. The iris image is thus transformed into a one bit template that is the same size as the image in polar coordinates (120 rows by 180 columns). In some portions of the image input to the filters, the energy may be too low to reliably determine if a ridge is present. For this reason, each template is accompanied by a binary mask, with a 1 indicating presence of a ridge and a 0 indicating no ridge being detected.[3] For future implementations of the RED algorithm, detection of eyelids, eyelashes and specularities will be incorporated into the segmentation, so that the mask will also be used to identify these non-iris areas as well as iris regions without prominent ridges. The template generation process is outlined in Fig. 6.

C. Cordic Modul

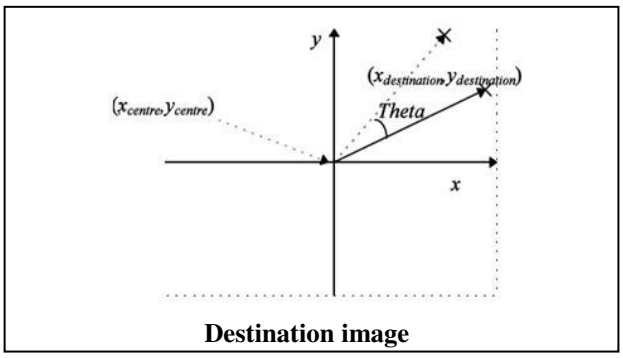


Figure 5: The points created by the CORDIC modules in one clock cycle

Creating a Circle with a Given Center-point and Radius. Given an (*x, y*) coordinate from the search area and an even radius between 10 and 126 pixels, a list of address points for a circle with center-point (*x, y*) and radius *r is created*. The circle plot area is a 256 by 256 pixel space taken from the edge detected image. The address points are computed by implementing a version of the CORDIC algorithm.[1] The CORDIC module is copied 91 times in parallel to create the first 91 points on a circle from t=0 to t=90 in one clock cycle.

$$X_0 = X_{center} + r\cos(t) \qquad (1a)$$

$$Y_0 = Y_{center} + r\cos(t) \qquad (1b)$$

In the memory controller these points will be used to form the full circle. After the 91 points are created they go into an array and are compared to remove any repeat points created due to rounding. Once this is complete, the points are ready to move into the memory controller. The memory controller takes three points at a time and rotates them into the second, third and fourth addresses for simultaneous reads. Each pixel location (*x1, y1*) in the first quadrant is used to generate the corresponding three coordinates (*x2, y2*), (*x3, y3*), (*x4, y4*) in the second, third and fourth quadrants.

Every time the memory controller fetches three data points from the CORDIC module, it accesses 12 memory locations in one clock cycle. To support this processing rate, six copies of the edge image are stored and dedicated to each CORDIC module. The 12 bits are read and sent to the third part of the architecture, the accumulator.

D. Template Matching

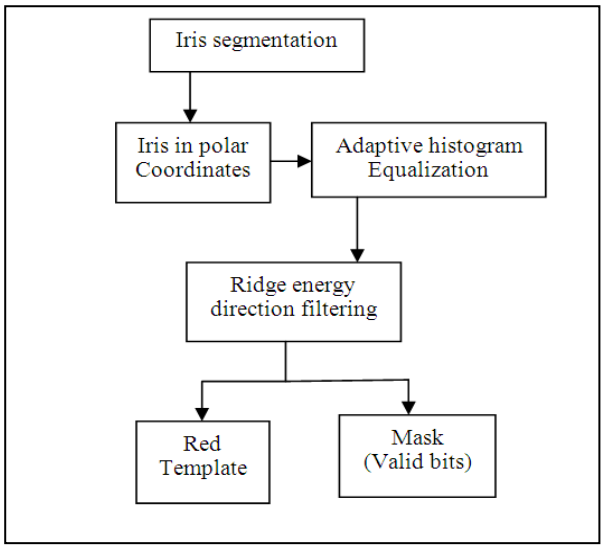


Figure 6: Template Matching

For matching, this template can now be compared to a stored template using fractional Hamming distance (HD) as the measure of closeness (template A template B) mask A mask B

$$HD = \frac{\|(\text{template A} \otimes \text{template B}) \cap \text{mask A} \cap \text{mask B}\|}{\|\text{mask A} \cap \text{mask B}\|}$$

The operator is the binary exclusive-or operation to detect disagreement between the bits that represent the directions in the two templates, is the binary AND function, ||.|| is a summation, and masks A and B are the associated binary masks for each template. [5] The denominator ensures that only the bits that matter are included in the calculation, after non-ridge areas are discounted. Rotation mismatch between irises (due to head-tilt) is handled with left right shifts of the template to determine the minimum HD. [3]For example, with 120 x 180 templates, each column represents of angular resolution and a shift of 12°(6 columns) is performed in each direction (left/right). The resulting fractional Hamming distances (similarity scores) representing genuine matches (i.e., comparisons of the same eye) and imposter matches (comparisons of different eyes) generated the results presented in the next section.

III. FPGA IMPLEMENTATION

A. Parallelizing iris recognition

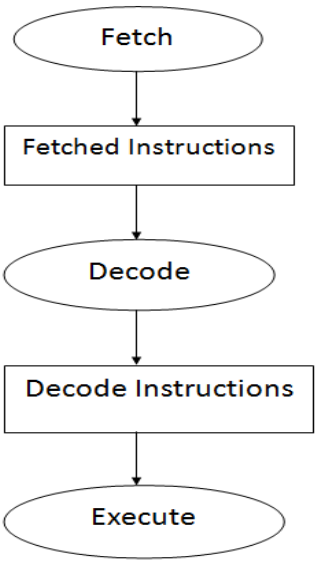


Figure 7: Simple processor pipeline with four ALUs

The iris recognition algorithms are currently implemented on general purpose sequential processing systems, such as generic central processing units (CPUs). In this work, we present a more direct and

parallel processing alternative using field-programmable gate arrays (FPGAs), offering an opportunity to increase speed and potentially alter the form factor of the resulting system. Within the means of this project, the most time-consuming operations of a modern iris recognition algorithm are deconstructed and directly parallelized. In particular, portions of iris segmentation, template creation, and template matching are parallelized on an FPGA-based system, with a demonstrated speedup of 9.6, 324, and 19 times, respectively, when compared to a state-of-the-art CPU-based version.

Figure 8 illustrates a simple processor pipeline. Instructions are first fetched, decoded, and finally executed by more than one ALU. Therefore, multiple instructions can be executed in parallel. However, modern processors are limited in the number of ALUs they possess, and most of today's CPUs do not have more than four logic units or ALUs. State-of-the-art processors contain more than one ALU. Therefore, multiple instructions can be executed in parallel. The focus of this research is on parallelizing key portions of the iris recognition algorithm using an FPGA. This work demonstrates this by making the following contributions, Introduction and calculation of the theoretical best performance of CPU-based machines executing key components of an iris recognition algorithm.

Measurements of CPU performance of key components of an iris recognition algorithm on a state-of-the-art computer. Deconstruction and novel parallelization of three key iris recognition components, Iris segmentation using local kurtosis, Iris template creation via filtering, and Template matching via hamming distance Evaluation of the benefits of parallelization in terms of performance and size.

After twelve points are read from memory they are sent directly into the accumulator. In one clock cycle, the 12 fell on the created circle. As the memory controller reads the next three points from the list of 91, rotates them and sends the addresses into memory, the accumulator adds the previous cycle's data. Once all points have been read and added together a flag indicates the circle has been read and a percentage of the circle is calculated. [1] A best percentage is stored in the module, each new percentage being compared to the current best. Each of the 36 parallel paths keeps its own best percentage. Once all center points and radii have been tested, the best of each module goes through a tree of comparators and the overall best match becomes the output of the system. The center-point, radius and percent match are output. This ends the process for the given image. The overall layout is shown in Fig. 8 for the parallel path.

B. Architecture

We utilize Field-Programmable Gate Arrays (FPGAs) to parallelize directional filtering. FPGAs are complex programmable logic devices that are essentially a "blank slate" integrated circuit from the manufacturer and can be programmed with nearly any parallel logic function. They are fully customizable and the designer can prototype, simulate and implement a parallel logic function without the costly process of having a new integrated circuit manufactured from scratch. In Figure 8 an FPGA [3] has been programmed with many ALUs. FPGAs are commonly programmed via VHDL (VHSIC Hardware Description Language). VHDL statements are inherently parallel, not sequential. VHDL allows the programmer to dictate the type of hardware that is synthesized on an FPGA. For example, if you would like to have 2,048 XOR logic gates that execute in parallel, then you program this directly in the VHDL code

Our results were performed on a modest-sized Spartan-3 FPGA, and we expect future generations of FPGAs to increase in performance greater than future generation CPU chips. We also believe other components of iris recognition can be ported to a parallel system. Iris recognition is becoming more and more popular, and an accurate, timely system is our goal.

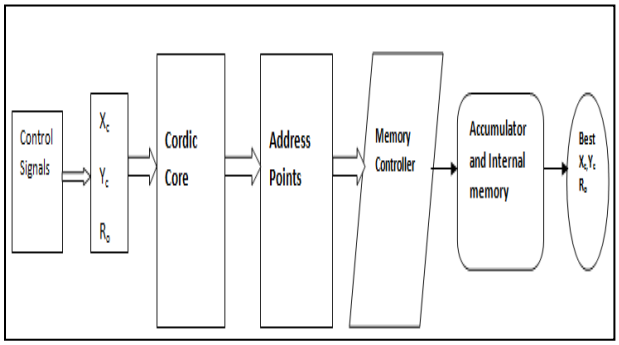


Figure: 8 layout of the parallel implementation of algorithm

IV. RESULTS AND ANALYSIS

A. A. Hardware Details

The CPU experiment is executed on an Intel Xeon X5355 workstation class machine. The processor is equipped with eight cores, 2.66-GHz clock, and an 8-MB L2 cache. While there are eight cores available, only one core is used to perform this test, therefore allowing all cache and memory resources for the code under test. The HD code was compiled under Windows XP using the Visual Studio software suite. The code has been fully optimized to enhance performance. Additionally, millions of matches were executed to ensure that the templates are fully cached in the on-chip

L2 cache. The RED algorithm optimized C++ code time is faster than some of the times reported in the literature for other commercial iris recognition implementations. For example, the template matching time earlier was 10m, whereas our template matching time is 383 ns. We attribute this difference to: 1) our use of a faster, more modern CPU operating at 2.66 GHz vice 300 MHz; 2) the RED algorithm is fully optimized;

	Percentage of logic elements on cyclic	Percentage of logic elements estimated on stratix	comment
Kurtosis	9%	.9%	Utilizes faster on-chip multipliers. After multipliers are exhausted, each parallel instantiation of kurtosis consumes 9%.
Digital filtering	71%	7.1%	This algorithm is hardware intensive. It also consumes 17% of on chip memory.
Hamming distance	73%	7.3%	Storing a template consumes 0.7% of cyclone memory. Therefore approximately 230 irises can be stored on cyclone. A tenfold increase can be stored on the stratix.

Figure 9: FPGA Hardware Usage Results

An extrapolation of the execution times of on a modern computing environment at 2.66 GHz would leave it still slower, but closer to the RED execution times. Our goal here is not to provide a direct comparison of the execution times of these two iris recognition algorithms. chip is necessary to execute our algorithm. We are able to determine the size required of our program on the FPGA, and the resulting execution time.

B. Comparisons

Speedup: All VHDL code is fully synthesizable and is downloaded onto our FPGA for direct hardware execution. As discussed above, our code is fully contained within a "process" statement. The process statement is only initiated when a signal in its sensitivity list changes values. The sensitivity list of our process contains the clock signal and, therefore, the code is executed once per clock cycle. In this code, the clock signal is drawn from our FPGA board which contains a 50-MHz clock. Therefore, every 20 ns, our calculation is computed. Note that for template generation, this process takes many clock cycles, and therefore, the total time for template generation with our FPGA is 98.6 us. Fig. 9 illustrates the execution times and acceleration achieved for our implemented FPGA.For example, the optimized C++ version takes 96.8 ns per match while the FPGA takes 20 ns per match for the kurtosis function.

The parallel algorithm on our FPGA is approximately 2.3, 4.7, and 7.5 times calculations, respectively. If Intel were to design a much faster microprocessor with a perfect compiler, it still would be orders of magnitude slower than an off-the-shelf inexpensive low-end FPGA.

V. CONCLUSION

The hardware implementation of RED is still in progress. This segment of the implementation of the RED algorithm can be improved to maximize clock frequency by adjusting the pipeline. The next implementation will include the Canny edge detection algorithm on the FPGA as well. [2]We also expect to find a speed-up on this part. There are other portions of RED that could be improved using a hardware design. Some of these include other parts of the segmentation process as well as the tedious bit-matching algorithm for finding a match within a large database.

This method carries with it several assumptions. First, it is assumed that the iris images are orthogonal, such that the eye is looking directly at the camera. In conjunction with this, it is assumed that the pupil and the limbic boundary of the iris is circular, which is not always accurate. The eyes are assumed to be wide open in that the presence of eyelids or eyelashes within the determined boundaries of the iris is not considered when extracting the ridge features, which serves to detriment the system performance. Overall, the algorithm has several areas that can be addressed to improve performance; these are discussed in the next section. Considerable gains in performance are expected with these modifications Iris template generation with directional filtering can be computationally intensive, yet the algorithms are currently designed for traditional sequential processing elements, such as a personal computer. We have presented a parallel processing alternative using field programmable gate arrays to speed up template generation. [5]

Iris recognition as a biometric technology has great advantage such as variability, stability and security. Thus it will have a variety of application. Here an image is analyzed by calculation of histogram and then it is converted to binary image for that purpose a reasonable threshold value is chosen. And then edge detection is done using a canny edge detector. The Canny's edge detection algorithm performs better than all other operators under almost all scenarios.

REFERENCE :

[1] Using an FPGA to Accelerate Pupil Isolation in Iris Recognition Jennifer L. Shafer, Hau Ngo, and Robert W. Ives United States Naval

Academy Department of Electrical and Computer Engineering Annapolis, MD 21402 USA

[2] J. Daugman, "Probing the uniqueness and randomness of IrisCodes: Results from 200 billion iris pair comparisons." *Proceedings of the IEEE*, vol. 94, no. 11, pp 1927-1935.

[3] Robert W. Ives, Randy Broussard, Lauren Kennell, Ryan Rakvic and Delores Etter, "Iris Recognition Using the Ridge Energy Direction (RED) Algorithm," *Proceedings of the 42nd Annual Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA,Nov. 2008.

[4] John Canny, "A Computational Approach to Edge Detection," *IEEE Transactions of Pattern Analysis and Machine Intelligence,* vol. PAMI-8, no. 6,

[5] Ryan N. Rakvic, Bradley J. Ulis, Randy P. Broussard, and Robert W. Ives, "Iris Template Generation with Parallel Logic," *Proceedings of the 42nd Annual Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2008.

[6] A Distributed Canny Edge Detector And Its Implementation On Fpga *Qian Xu, Chaitali Chakrabarti and Lina J. Karam* School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ qianxu@asu.edu, chaitali@asu.edu, karam @asu.edu

[7] W. He and K. Yuan, "An improved Canny edge detector and its realization on FPGA," *WCICA*, pp. 6561 –6564, Jun. 2008.

[8] Monro, D. M., Rakshit, S., and Zhang, D, University of Bath, U.K. Iris Image Database, http://www.bath.ac.uk/ elec eng/pages/sipg/ irisweb

[9] A fast method for iris localization R. Meenakshi Sundaram Dept. of Information Technology Jadavpur University.2011

[10] Altera, "TriMatrix Embedded Memory Blocks in Stratix IV Devices," *Alter Corporation Stratix IV Device Handbook*, vol 1, March 2010.

◈◈◈

Energy Efficient Routing Protocol for Dynamic Nodes in WSN

Robin Victor J, Mani Mozhi I. & Jitendranath Mungara

Dept of CSE, CMR Institute of Technology, Bangalore
E-mail : robinvictor87@gmail.com, srimanisen@gmail.com, jmungara@yahoo.com

Abstract **-** Wireless sensor network consists of many tiny disposable, low power devices with sensing, processing and wireless communication capabilities. The design of routing protocol for wireless sensor network is influenced by various factors such as data aggregation, network heterogeneity, energy efficiency, mobility of nodes, scalability, etc. Since each sensor node has a finite energy supply and bandwidth, the primary objective of the design is to enhance the energy efficiency. Most of the existing routing protocols for WSN have been designed for static nodes, but however this assumption is invalid for most application. To improve the energy efficiency and to support dynamic nature of the node, in this paper we introduce a new hybrid routing protocol called 'HAODV' (Hybrid Ad hoc On –Demand Distance Vector) which is based on the characteristics of Threshold - sensitive Energy Efficient sensor Network protocol and Ad hoc On-demand Distance Vector, in order to allow the nodes to move dynamically we make use of Random walk mobility model and further to enhance the energy efficiency we adopt a new technique called Random-Cast protocol.

Keywords **-** *AODV, TEEN, Wireless Sensor Networks, Energy efficiency, HAODV, dynamic nodes, Random-Cast.*

I. INTRODUCTION

Wireless Sensor Networks can be generally described as a network of nodes that cooperatively sense and may control the environment enabling interaction between persons or computers and the surrounding environment.

A WSN consists of large number of sensors and one or more sinks where data is collected. Each sensor node is able to sense the physical environment, process data locally, and participates in data forwarding to a sink, from where data are retrieved by users. WSNs are application specific, thus sensors nodes are equipped with sensors accordingly. Some applications (e.g. building monitoring) require a smaller number of sensors that can be placed individually. Others (e.g. surveillance of a battlefield) require a large number of sensors (e.g. thousands or even millions) that will be deployed ad hoc .Using a larger number of sensors increases network robustness and fault-tolerance. Sensor nodes can be organized in a wireless sensor network (WSN). The main functions of a sensor node are data sensing, local data processing, and data forwarding. Figure 1 shows the general architecture of a WSN.

WSNs can be classified as homogeneous and heterogeneous. Homogeneous WSNs contain only one type of devices, the sensor nodes. On the other hand, heterogeneous WSNs contain devices of different capabilities. Typical sensor nodes are resource-constrained nodes and another type, called Base station here, are more resource rich than the standard sensor nodes; they can have, for example, more energy resources, larger transmission range, higher data rate, etc.

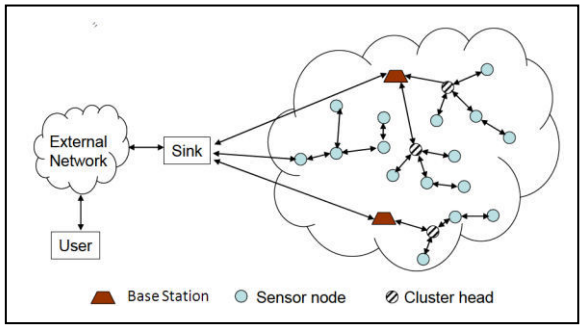


Fig. 1: Architecture of a WSN

Energy strictly limited and dynamic topologies are two characteristics in wireless sensor networks. Because of energy limited, it's key to achieve best balance between reliable data transmission and energy consumption; taking dynamic topology into account, how to maintain a real-time network by minimum cost is important.

Mobility in wireless sensor networks (WSN) has attracted a lot of attention in the recent years. Most of the existing protocols assume that the nodes are stationary, since this assumption facilitates the simplification of the protocols, making them have a very low overhead. However, in applications like elderly health monitoring, habitat monitoring, search and rescue, this assumption makes those clustering

mechanisms invalid, since the static nature of sensors is not real for these applications. The issues related to physical and link layers are unique for all kind of sensor applications, therefore the research on these areas has been focused on system-level power awareness such as dynamic voltage scaling, radio communication hardware, low duty cycle issues, system partitioning, energy aware MAC protocols[2].

In this paper, reactive routing protocols Ad hoc On-demand Distance Vector (AODV) and Threshold - sensitive Energy Efficient sensor Network protocol (TEEN) are analyzed based on their strengths, attributes, perspective of design objectives, assumptions, and drawbacks. AODV establishes a route from one node to another node when there is a need. This is one of the important advantages of AODV which makes it viable for use it as a routing protocol in WSNs [5, 6].AODV fails in energy management. Threshold sensitive Energy Efficient sensor Network protocol which is energy efficient using threshold constraints [6], fails in finding the active nodes and maintain the route from node to base station.

Based on the analysis of two reactive routing protocols namely AODV (Ad hoc on- demand Distance Vector) and TEEN (Threshold - sensitive Energy Efficient sensor Network Protocol) a new hybrid routing protocol called HAODV is used which incorporates the constraints of TEEN in AODV. The proposed routing protocol HAODV is significantly better than traditional AODV. This protocol finds active nodes and maintain the route while transmission of data which is better than TEEN.

Some WSN routing protocols such as Ad hoc On-demand Distance Vector (AODV) collect route information via overhearing, they would suffer if they are used in combination with 802.11 PSM. Allowing no overhearing may critically deteriorate the performance of the underlying routing protocol, while unconditional overhearing may offset the advantage of using PSM. This paper proposes a new communication mechanism, called RandomCast, making a prudent balance between energy and routing performance. In addition, it reduces redundant rebroadcasts for a broadcast packet, and thus, saves more energy. Therefore, to enhance the overall network performance we also make use of Randomized overhearing using Random-Cast which mainly improves the energy performance by controlling the level of overhearing and forwarding without causing significant impact on the network performance.

The remainder of this paper is organized as follows. Section II discusses some related work. Section III describes the proposed algorithms,. Finally, Section IV presents our conclusions and future directions.

II. RELATED WORK

A. AODV (Ad hoc On-demand Distance Vector):

AODV is a reactive routing protocol and a route from one node to another node can be found when required [2]. Each intermediate node in the network forwards the Route Request (RREQ) message until it reaches the destination node. The destination node responds to the RREQ message by transmitting the Route Reply (RREP) message. As the RREP flows through the network, it determines the route from source node to destination node. The sequence number is increased by each originating node and used to determine whether the received message is the most recent one [6]. The older routing table entries are replaced by the newer ones. Active nodes in the networks are determined by broadcasting a "Hello" message periodically in the network. If a node fails to reply a link break is detected and a Route Error (RERR) message is transmitted which is used to invalidate the route as it flows through the network. A node also generates a RERR message if it gets message destined to a node for which a route is unavailable.

Types of messages in AODV:

1. Route Request (RREQ) message: It is used to form a route from one node to another node in a network.
2. Route Reply (RREP) message: It is used to connect destination node to source node in a network.
3. Route Error (RERR) message: It is used to indicate any route broken or node failure.
4. HELLO message: It is used to determine the activeness of the network.

B. Threshold - sensitive Energy Efficient sensor Network protocol (TEEN):

TEEN is a reactive protocol proposed for time-critical applications [6]. BASE STATION broadcasts the attribute, Hard Threshold (HT) and Soft Threshold (ST) values to its cluster members. The sensor nodes starts sensing and transmits the sensed data when it exceeds HT. HT is the minimum attribute range above which the values are expected. The transmitted sense value is stored in an internal variable called Sensed Value (SV). The cluster nodes again starts sensing, when its value exceeds the ST. The minimum change in the sensed value it switches on its transmitter and transmits.

The energy is conserved since the sensor nodes in the cluster senses continuously but transmits only when the sensed value is above HT. The ST further reduces the transmission which could have been occurred when there is a little change or no change in sensed attribute. As the cluster heads (CH) need to perform extra

computations it consumes more energy compared to other nodes. The main drawback of this protocol is that the transmission from nodes to CH will not be there when the sensed value is not greater than HT, hence the CH will never come to know even when any one of the sensor node dies. Accurate and clear picture of the network can be obtained by fixing the ST as smaller value even though it consumes more energy due to frequent transmissions [6].

C. Random Walk Model

The Random Walk model was originally proposed to emulate the unpredictable movement of particles in physics. It is also referred to as the Brownian motion. Because some mobile nodes are believed to move in an unexpected way, Random Walk mobility model is proposed to mimic their movement behavior [7].

In the Random Walk model, the nodes change their speed and direction at each time interval. For every new interval t, each node randomly and uniformly chooses its new direction θ (t) from [0, π2]. In similar way, the new speed follows a uniform distribution or a Gaussian distribution from [0, V]. Therefore, during time interval t, the node moves with the velocity vector [v(t)cosθ(t), v(t)sinθ(t)]. If the node moves according to the above rules and reaches the boundary of simulation field, the leaving node is bounced back to the simulation field with the angle of θ (t) or π-θ (t) respectively. This effect is called border effect.

III. PROPOSED ALGORITHM

A. Assumptions:

1) In the proposed model, nodes move randomly, after deployment, maintaining a dynamic topology based on Random walk mobility model.

2) The number of clusters and Cluster Head formed is dynamic. As a cluster member wants to transmit data to base station, it transmits the data through the cluster head.

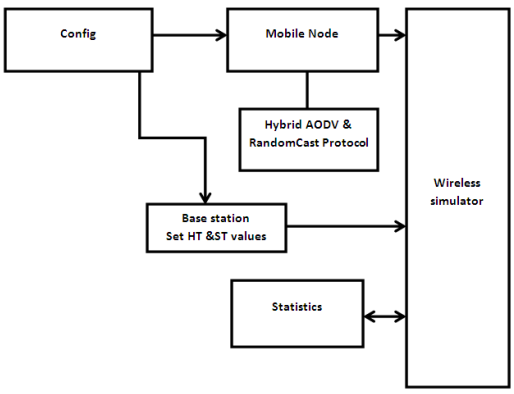


Fig. 2: System Architecture

Figure 2 shows the System Architecture for the proposed design, where the user will configure the system using the config module. The configuration will be in terms of number of nodes, mobility pattern of nodes, initial energy level, etc. Mobile Node and Base station communicate through the wireless simulator. Statistics module gathers the result from the wireless simulator like packet delivery ratio, threshold and plot in a neat graph. Mobile Node use Hybrid AODV protocol for routing messages.

i.) HAODV Protocol:

The constraints of TEEN are incorporated in AODV routing protocol. In the modified algorithm each cluster node sense the data, if the value is greater than HT, then node sends RREQ to the destination. By receiving RREP message from destination source node transmits the data to destination node. The HT value is stored in SV a variable which stores the transmitted threshold value. The cluster nodes again starts sensing, when its value exceeds the ST i.e. The minimum change in the sensed value occurs it switches on its transmitter and transmits. . Active nodes in the networks are determined by broadcasting a "Hello" message periodically in the network. If a node fails to reply a link break is detected and a Route Error (RERR) message is transmitted which is used to invalidate the route as it flows through the network. A node also generates a RERR message if it gets message destined to a node for which a route is unavailable.

As the properties of AODV and TEEN are bestowed in the modified hybrid algorithm (HAODV), better results are obtained is related to energy consumption than AODV and it can find the active nodes which is not possible with TEEN.

ii.) Random Cast Protocol:

The proposed Random Cast protocol is designed to improve energy performance by controlling the level of overhearing and forwarding without a significant impact on network performance. The main advantage of the proposed scheme is that the nodes are consistently operated in the PS mode. In Random Cast, a transmitter is able to specify the desired level of overhearing. The Randomized overhearing that chose not to overhear will switch to a low-power sleep state during the following data transmission period, saving substantial amount of energy compared to unconditional overhearing. With respect to route information, this does not necessarily deteriorate the quality of route information due to its spatial and temporal locality of route information. The Random Cast algorithm can be applied to broadcast packets such as RREQ to allow randomized rebroadcast. This is to avoid redundant rebroadcasts of the same packet in the mobile networks. On the other hand, the

rebroadcast decision must be made conservatively. This is because a broadcast packet may not be delivered to all nodes in the network when randomized rebroadcast is used. For example, an RREQ packet may not reach the specific destination node. For this reason, rebroadcast probability (PF) is set higher than overhearing probability (PR).The Random Cast implementation is randomization. Basically, each node maintains an overhearing (rebroadcast) probability, PR (PF), determined using the factors sender ID, number of neighbors. Mobility and remaining battery energy. In other words, if a randomly generated number is higher than PR, then a node decides to overhear (rebroadcast).

The overall dataflow diagram for the proposed design is as shown in figure 3.

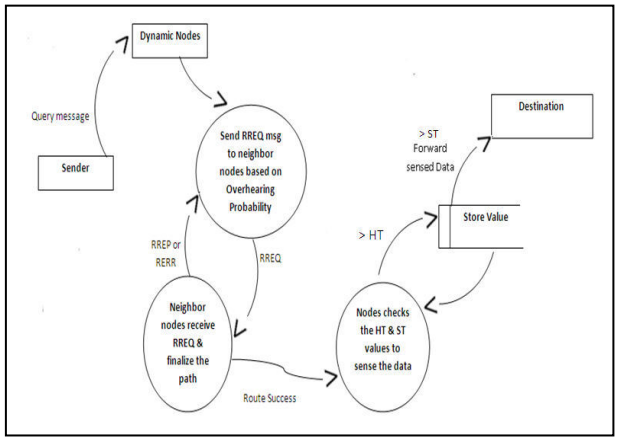


Fig. 3: Dataflow diagram for the overall design

Parameters used in network:

- Channel: Wireless
- MAC protocol: 802.11/RandomCast
- Antenna: Omni directional
- Energy model: Battery
- Initial energy: 10J
- Number of nodes: 50
- Mobility of nodes: Dynamic
- Number of clusters: 8
- Data Packet size: 512bps

The input for the graphics is stated above. The simulation tool used is NS 2.32. It provides substantial support for simulation of TCP, routing and multicast protocols. The simulator is written in C++ and script language is OTcl. The user writes the script in OTcl to define a network (nodes and links), the traffic in the network (source, destination and type of traffic) and which protocols it will use. The results of the simulations are an output trace file that can be used to data processing and to visualize simulations with a program called Network Animator (NAM) [10, 11].

IV. CONCLUSION

The energy efficient hybrid HAODV protocol supporting dynamic nodes is presented in this paper. This algorithm employs the energy-aware strategy for data communication using RandomCast protocol and also it improves the overall network performance such as throughput, end-to-end delay, packet delivery ratio. The proposed system is well suited for continuous model of dynamically distributed Wireless Sensor Networks. The proposed protocol will perform well in terms of energy efficiency, throughput, packet delivery ratio and end-to-end delay when compared to other routing protocol. In future, topological transformation for several nodes will be conducted so as to improve the algorithm to suit for highly complex distributed WSNs.

REFERENCES

[1] Chiara Buratti, Andrea Conti, Davide Dardari and Roberto Verdone “An Overview on Wireless Sensor Networks Technology and Evolution”, Sensors, PP: 6869-6896, 2009.

[2] Akyildiz, I.; Su, W.; Sankarasubramaniam, Y.; Cayirci, E. “A survey on sensor networks” IEEE Commun., Vol: 40, PP:102–114, March. 2002.

[3] Kemal Akkaya and Mohamed Younis “ Survey on Routing Protocols for Wireless Sensor Networks” ,IEEE,2005.

[4] J. Chang and L. Tassiulas, “Maximum lifetime routing in wireless sensor networks,” IEEE/ACM Transactions on Networking (TON), vol. 12, no. 4, pp. 609–619, 2004.

[5] I. Chakeres and E. Belding-Royer, “AODV routing protocol implementation design,”. Proceedings. 24th International Conference on in Distributed Computing Systems Workshops, pp. 698–703,2004.

[6] M.Dhivya, Ashok Kumar Potnuru, M.Sundarambal “Energy Efficient Routing Protocol for Wireless Sensor Networks”. www.ieee.org Vol. 978-1-61284-764-1/11/2011 IEEE.

[7] Fan Bai and Ahmed Helmy “A Survey of Mobility Models in Wireless Ad hoc Networks” - , University of Southern California, U.S.A 2009.

[8] D.Baghyalakshmi, Jemimah Ebenezer and S.A.V .Satyamurty, “Low Latency And Energy Efficient Routing Protocols For Wireless Sensor Networks” International Conference on Wireless

Communication and Sensor Computing, ICWCSC IEEE FEB 2010.

[9] Sunho Lim, Chansu Yu, and Chita R. Das, "RandomCast: An Energy-Efficient Communication Scheme for Mobile Ad Hoc Networks",IEEE TRANSACTIONS ON MOBILE COMPUTING, VOL. 8, NO. 8, AUGUST 2009.

[10] Teerawat Issariyakul and Ekram Hossain "Introduction to Network Simulator NS2" 2009. [11] Teerawat Issariyakul and Ekram Hossain "Introduction to Network Simulator NS2" 2009.

[11] De los Andes, Merida, venezuala and ESSI "NS simulator for beginners" Lecture notes: 2003-2004.

◈◈◈

Mobility Assisted - Ad Hoc On-demand Distance Vector Routing

Rajani P, Sridevi K N & Jitendranath Mungara

CSE Dept., CMRIT, Bangalore, India
E-mail : rajanibmrr@gmail.com, sridevi.cmrit@gmail.com

Abstract **-** The proposed work highlights a novel work being formulated to design framework for mobility in MANET for the purpose of minimizing packet loss n MANET. Packet loss due to frequent link breakage is a very significant issue in MANET due to its dynamic topology. The proposed system has used the potential of Ad Hoc on-demand Distance Vector (AODV) which considers frequent transmission failures as well as power factor for preventing packet loss and enhancing the throughput. The proposed protocol evaluates the quantity of data packet which can be communicated through the transmission link prior it breaks before the transmission takes the place in MANET scenario. Simulation is performed in java environment to show that the delivery ratio can be improvised extensively in mobile ad hoc network.

Keywords **-** *component; Mobility, Mobile ad hoc network, packet loss, AODV.*

I. INTRODUCTION

In general, mobile ad hoc networks [1] are formed dynamically by an autonomous system of mobile nodes that are connected via wireless links without using the existing network infrastructure or centralized administration. The nodes are free to move randomly and organize themselves arbitrarily; thus, the network's wireless topology may change rapidly and unpredictably. Such a network may operate in a standalone fashion, or may be connected to the larger Internet. Wireless mobile ad hoc networks (MANETs), due to their dynamic nature, pose many unique challenges compared to traditional wired or cellular wireless networks. Mobile ad hoc networks are infrastructure-less networks since they do not require any fixed infrastructure, such as a base station, for their operation. In general, routes between nodes in an ad hoc network may include multiple hops, and hence it is appropriate to call such networks as ''multi-hop wireless ad hoc networks''. Each node will be able to communicate directly with any other node that resides within its transmission range. For communicating with nodes that reside beyond this range, the node needs to use intermediate nodes to relay the messages hop by hop. In this project work, we focus on presenting an algorithm to include mobility in a routing protocol to reduce packet losses in mobile ad hoc networks (MANETs). The algorithm is applicable to any on-demand routing protocol. The proposed algorithm estimates the number of packets that can traverse through the route before it breaks because of mobility. This algorithm increases network throughput and packet delivery ratio. Our proposed algorithm estimates the route's RLT, and allows routing protocol to send number of packets that can traverse the route successfully during the RLT period. This provision is implemented and shown effective in improving the routing performance in MANETs. Packet loss occurs when one or more packets of data travelling across a computer network fail to reach their destination. Packet loss is distinguished as one of the three main error types encountered in digital communications; the other two being bit error and spurious packets caused due to noise. Packet loss can be caused by a number of factors including signal degradation over the network medium due to multi-path fading, packet drop because of channel congestion, corrupted packets rejected in-transit, faulty networking hardware, faulty network drivers or normal routing routines (such as DSR in ad-hoc networks). Mobile ad hoc networks with 10, 20 and 30 nodes are considered for testing the proposed algorithm. Each node in this network is considered to be equipped with GPS (or any other positioning service) which is capable of providing the current location, direction of movement, as well as the current velocity of the mobile node. Packet losses occur in two ways (i.e. due to link failures and due to collisions), and the emphasis here is given to reduce packet losses due to link failures. We have created two routes from source to destination that can be broken and reconnected continuously. For the complete functionality of the project work, the project is run with the help of well equipped computer containing at least P4 processor, 20 GB HDD and 2 GB RAM. Normally, the OS is Windows or Linux. The main theme of this project work is to introduce an algorithm to include mobility in a routing protocol to reduce

packet losses in MANET. It also estimate the number of packets in mobile ad hoc networks (MANETs) that can traverse through the route before it breaks because of mobility. This is achieved by viewing the different node activities inside the network and the specified way has to be mentioned to monitor those activities. Performance analysis will be done in order to estimate the number of packets in MANET among all nodes. In section 2 we give an overview of related work which identifies all the major research work being done in this area. Section 3 highlights problem descriptions. Proposed system is discussed in Section 4 followed by implementation and analysis in Section 5. Section 6 discusses some concluding remarks.

II. RELATED WORK

Djamel Djenouri et.al [2] drive a GloMoSim based simulation study, to investigate the mobility effects on the performance of several mobile ad hoc routing protocols.

Feng Li and Jie Wu [3] present a certainty oriented reputation system that emphasizes the relationship among uncertainty, observation and recommendation

Bhavyesh Divecha et.al [4] has studied the effects of various mobility models on the performance of two routing protocols Dynamic Source Routing (DSR-Reactive Protocol) and Destination-Sequenced Distance-Vector (DSDV-Proactive Protocol).

Andrea E.F. Clementi et.al [5] consider a Mobile Ad-hoc NETworks (MANET) formed by n nodes that move independently at random over a finite square region of the plane

Nils Aschenbruck et.al [6] presents a realistic approach to realize the mobility in disaster areas based on tactical issues of civil protection.

R. Timo et.al [7] studies the stability of a class of discrete Random Waypoint Mobility Models (RWMMs). This class includes the classic Random Waypoint Mobility Model.

Xianren Wu et.al [8] present statistical models to accurately evaluate the distribution of the lifetime of a wireless link in a mobile ad hoc network (MANET) in which nodes move randomly within constrained areas.

Yasser Kamal Hassan et.al [9] presented a comparing performance of protocols for routing packets between wireless mobile hosts in an ad-hoc network AODV, DSR and DSDV using a network simulator like NS-2 with scenario consist of dynamic network size and di®erent number of movement speed at invariable pause time which used an AODV and DSR from On-Demand protocols compared with DSDV from proactive table-driven routing protocols.

Baldev Ram Mali and N.C. Barwar [10] carried out to compare the performance of one proactive routing protocol DSDV and two reactive protocols AODV and DSR of MANETs under both CBR and TCP traffic patterns using NS-2 simulator.

Gang Lu et.al [11] propose a novel Environment-Aware Mobility (EAM) model which models a more realistic movement of mobile nodes.

Juan-carlos cano et.al [12] presents an analysis of the effect that mobility models have over the performance of a mobile ad-hoc networks. They proposed four different group mobility models and describe a mobility pattern generator called grcmob to be used with the ns2 simulator.

Theus Hossmann [13] presents an approach to mobility prediction based on pattern matching.

Xiaofeng Lu et.al [14] analyze a MANET military scenario traces in the structural, spatial and temporal characteristics.

Mr. N. Kumaresh [15] describes a distributed mobility management scheme for mobile sensor networks.

Suprio Ray [16] introduces GEMM, a tool for generating mobility models that are both realistic and heterogeneous.

Illya Stepanov et.al [17] describes a comprehensive and extensible approach to model mobility of users in outdoor scenarios.

Maoning Wang [18] a novel scheme to integrate MANETs with the Internet and support mobility across Wireless Local Area Networks (WLANs) and MANETs.

III. PROBLEM DESCRIPTION

Regardless of the attractive applications, the features of MANET introduce several challenges that must be studied carefully before a wide commercial deployment can be expected. These include:

- Routing. Since the topology of the network is constantly changing, the issue of routing packets between any pair of nodes becomes a challenging task. Most protocols should be based on reactive routing instead of proactive. Multicast routing is another challenge because the multicast tree is no longer static due to the random movement of nodes within the network. Routes between nodes may potentially contain multiple hops, which is more complex than the single hop communication.
- Security and Reliability. In addition to the common vulnerabilities of wireless connection, an ad hoc

network has its particular security problems due to e.g. nasty neighbor relaying packets. The feature of distributed operation requires different schemes of authentication and key management. Further, wireless link characteristics introduce also reliability problems, because of the limited wireless transmission range, the broadcast nature of the wireless medium (e.g. hidden terminal problem), mobility-induced packet losses, and data transmission errors.

- Quality of Service (QoS). Providing different quality of service levels in a constantly changing environment will be a challenge. The inherent stochastic feature of communications quality in a MANET makes it difficult to offer fixed guarantees on the services offered to a device. An adaptive QoS must be implemented over the traditional resource reservation to support the multimedia services.
- Internetworking. In addition to the communication within an ad hoc network, internetworking between MANET and fixed networks (mainly IP based) is often expected in many cases. The coexistence of routing protocols in such a mobile device is a challenge for the harmonious mobility management.
- Power Consumption. For most of the light-weight mobile terminals, the communication-related functions should be optimized for lean power consumption. Conservation of power and power-aware routing must be taken into consideration.

Mobility is a crucial factor in MANETs and it plays an important role in determining the overall performance of the network this is because the high mobility of nodes can cause frequent changes in network topology, leading to less reliable routes and frequent link breakages, hence, increasing the re-initiation of the route discovery process, resulting in more control packets overhead due to the extra use of Route Request Packets (RREQ), Route Reply Packets (RREP), and Route Error Packets (RERR), and increasing the average end-to-end delay

IV. PROPOSED SYSTEM

The aim of the project work is to design an architectural framework for creating the environment of mobile ad hoc network to design mobility aware routing protocols in order to avoid the frequent link breakages associated with using unstable paths that contain high mobile nodes. The main objectives of the proposed research work can be briefed as following:

- To design a novel mobility assisted routing protocol using AODV for the MANET.
- To identify the node mobility.
- To reduce the link breakage.
- To avoid the end to end delay problem.
- The nodes perform mobility quantification in a distributed manner based on the locally available information about neighborhood nodes.
- To select the best neighbor node this has low mobility.
- To reducing reinitiating route discovery process and reduce the number of dropped packet

The problem of link failures in mobile ad hoc networks (MANETs) due to mobility is amplified when a route constitutes several such links. If any of those links fails, the route breaks, which initiates series of undesirable events and outcomes. If how long a link can be operational is predicted, the routing protocol can use this to its advantage. It is assumed that every link remains connected for a limited time, called link life time (LLT), and a route has a limited life, called route life time (RLT). The RLT depends on the LLTs of links, and can be taken as the lowest LLT in the route. When degree of mobility increases, LLTs and eventually RLTs decrease. That contributes to increase in packet losses and low throughputs in a MANET. Since mobility is inevitable in MANETs, an algorithm that utilizes mobility to decrease packet loss in the network is designed, and presented in our work. Our proposed algorithm estimates the route's RLT, and allows routing protocol to send number of packets that can traverse the route successfully during the RLT period. This provision is implemented and shown effective in improving the routing performance in MANETs. The proposed Architecture is as shown below:

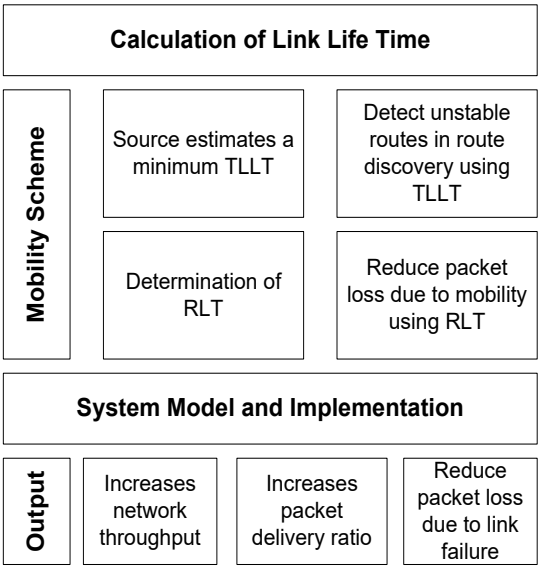


Fig.1. Proposed Architecture

A route discovered between a source node and destination node is maintained as long as needed by the source node. Since there is movement of nodes in mobile ad hoc network and if the source node moves during an active session, it can reinitiate route discovery mechanism to establish a new route to destination. Conversely, if the destination node or some intermediate node moves, the node upstream of the break initiates Route Error (RERR) message to the affected active upstream neighbors/nodes. Consequently, these nodes propagate the RERR to their predecessor nodes. This process continues until the source node is reached. When RERR is received by the source node, it can either stop sending the data or reinitiate the route discovery mechanism by sending a new RREQ message if the route is still required.

When the destination node or an intermediate node with a route to the destination receives the RREQ, it creates the RREP and unicast the same towards the source node using the node from which it received the RREQ as the next hop. When RREP is routed back along the reverse path and received by an intermediate node, it sets up a forward path entry to the destination in its routing table. When the RREP reaches the source node, it means a route from source to the destination has been established and the source node can begin the data transmission.

V. IMPLEMENTATION AND ANALYSIS

The entire proposed model is design in java environment on 32-bit Windows OS with 1.84 GHz processor speed. A node MAY offer connectivity information by broadcasting local Hello messages. A node SHOULD only use hello messages if it is part of an active route. Every HELLO_INTERVAL milliseconds, the node checks whether it has sent a broadcast (e.g., a RREQ or an appropriate layer 2 message) within the last HELLO_INTERVAL. If it has not, it MAY broadcast a RREP with TTL = 1, called a Hello message, with the RREP message fields set as follows:

- Destination IP Address: The node's IP address.
- Destination Sequence Number: The node's latest sequence number.
- Hop Count: 0
- Lifetime: ALLOWED_HELLO_LOSS * HELLO_INTERVAL

Parameter Name	Value
ACTIVE_ROUTE_TIMEOUT	3,000 Milliseconds
ALLOWED_HELLO_LOSS	2
BLACKLIST_TIMEOUT	RREQ_RETRIESxNET_ TRAVERSAL_TIME
HELLO_INTERVAL	1,000 Milliseconds
LOCAL_ADD_TTL	2
MAX_REPAIR_TTL	0.3 * NET_DIAMETER
MY_ROUTE_TIMEOUT	2*ACTIVE_ OUTE_TIMEOUT
NET_DIAMETER	35
NET_TRAVERSAL_TIME	2 * NODE_TRAVE RSAL _ TIME * NET_DIAMETER
NEXT_HOP_WAIT	NODE_TRAV ERSAL_TIME + 10
NODE_TRAVERSAL_TIME	40 milliseconds
PATH_DISCOVERY_TIME	2 * NET_TRAVE RSAL_TIME
RERR_RATELIMIT	10
RING_TRAVERSAL_TIME	2 * NODE_TRAV ERSAL_TIME * (TTL_VALUE + TIMEOUT_BUFFER)
RREQ_RETRIES	2
RREQ_RATELIMIT	10
TIMEOUT_BUFFER	2
TTL_START	1
TTL_INCREMENT	2
TTL_THRESHOLD	7

A node MAY determine connectivity by listening for packets from its set of neighbors. If, within the past DELETE_PERIOD, it has received a Hello message from a neighbor, and then for that neighbor does not receive any packets (Hello messages or otherwise) for more than ALLOWED_HELLO_LOSS * HELLO_INTERVAL milliseconds, the node SHOULD assume that the link to this neighbor is currently lost. Whenever a node receives a Hello message from a neighbor, the node SHOULD make sure that it has an active route to the neighbor, and create one if necessary. If a route already exists, then the Lifetime for the route should be increased, if necessary, to be at least ALLOWED_HELLO_LOSS * HELLO_INTERVAL. The route to the neighbor, if it exists, MUST subsequently contain the latest Destination Sequence Number from the Hello message. The current node can now begin using this route to forward data packets. Routes that are created by hello messages and not used by any other active routes will have empty precursor lists and would not trigger a RERR message if the

neighbor moves away and a neighbor timeout occur. This section gives default values for some important parameters associated with AODV protocol operations. A particular mobile node may wish to change certain of the parameters, in particular the NET_DIAMETER, MY_ROUTE_TIMEOUT, ALLOWED_HELLO_LOSS, RREQ_RETRIES, and possibly the HELLO_INTERVAL. In the latter case, the node should advertise the HELLO_INTERVAL in its Hello messages, by appending a Hello Interval Extension to the RREP message. Choice of these parameters may affect the performance of the protocol. Changing NODE_TRAVERSAL_TIME also changes the node's estimate of the NET_TRAVERSAL_TIME, and so can only be done with suitable knowledge about the behavior of other nodes in the ad hoc network. The configured value for MY_ROUTE_TIMEOUT MUST be at least 2 * PATH_DISCOVERY_TIME.

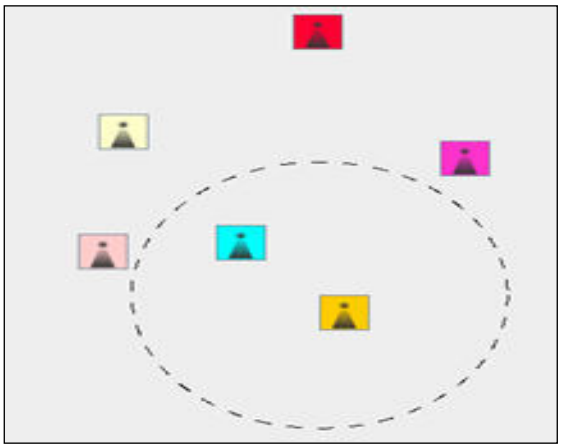


Fig.2. Mobility of the nodes

The application is initiated by creating a mobile node using IP address, name of node, position of node, and power of nodes. Once the simulation is initiated, the mobility of the different mobile nodes can be visualized as seen on Figure 2 above. The dotted line highlights the transmission range of nodes.

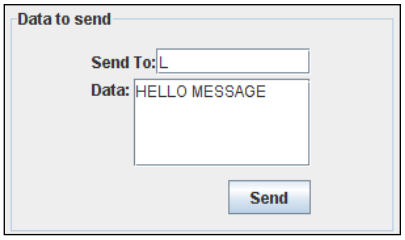


Fig.3. Inserting data for communication

Once the simulation is in progress, the source as well as destination node is selected by choosing the name of the node along with feeding certain data for communication as visualized in Figure 3.

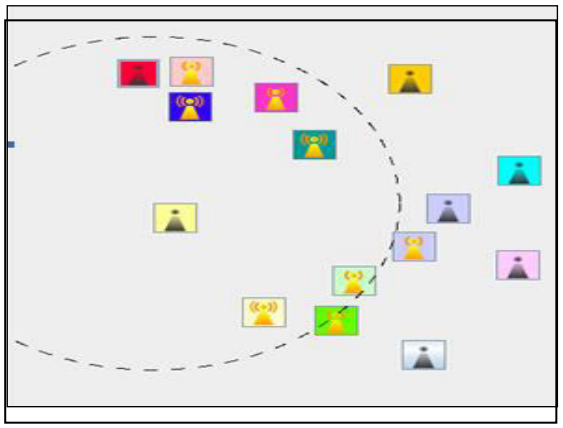


Fig.4. Message communication

The above figure 4 shows the process of communication as seen as yellow wireless icon. But, we cannot expect the communication instantly. As we have used random mobility, hence once the nodes come within the range of each other, then only we can expect an optimal communication.

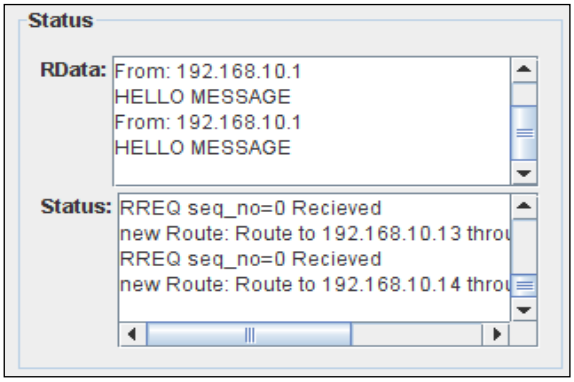


Fig.5. Insert Caption

It is not at all necessary that when the nodes come in the range of each other, it will initiate message communication. As due to random mobility, the intermediate transmitting node might wants to move very faster from their current location exactly in next instant. In that case, the reliability of the packet transmission cannot be evaluated. Hence the application is also equipped with evaluation of received data as well as status of the received data as seen on Figure 5.

VI. CONCLUSION

Mobile Ad Hoc Network (MANET) is collection of multi-hop wireless mobile nodes that communicate with each other without centralized control or established infrastructure. The wireless links in this network are highly error prone and can go down frequently due to mobility of nodes, interference and less infrastructure. Therefore, routing in MANET is a critical task due to highly dynamic environment. In recent years, several

routing protocols have been proposed for mobile ad hoc networks and prominent among them are DSR, AODV and TORA. This research paper provides an overview of these protocols by presenting their characteristics, functionality, benefits and limitations and then makes their comparative analysis so to analyze their performance. The objective is to make observations about how the performance of these protocols can be improved. A Mobile Ad hoc Network (MANET) is a collection of mobile nodes that communicate and collaborate with each other without reliance on any pre-existing infrastructure. In MANETs, wireless links are subject to frequent breakages due to nodes high mobility. While several routing protocols such AODV has been designed for MANETs, many of operate efficiently under low network mobility conditions and do not adapt well with high mobility conditions. Therefore, considering mobility is a demanding task that should be performed efficiently and accurately. In this paper, we propose novel mobility-aware routing protocols based on the well known Ad hoc On-demand Distance Vector (AODV) routing protocol called: MA-AODV (Mobility Aware-Ad Hoc On-Demand Distance Vector) in an attempt to improve the handling of high mobility in ad hoc networks. MA-AODV protocols perform periodic quantification of nodes mobility for the sake of establishing more stable paths between source/destination pairs, hence, avoiding the frequent link breakages associated with using unstable paths that contain high mobile nodes. Simulations are created using Java Swing in NetBeans. According to the results, our proposed protocols prove their superiority over the original AODV protocol in terms of the reduced overhead and the increased packet delivery ratio.

REFERENCES

[1] Jun-Zhao Sun , Mobile ad hoc networking: an essential technology for pervasive computing, Info-tech and Info-net, Proceedings, Beijing, International Conferences, 2001

[2] Djamel Djenouri, Abdelouahid Derhab, and Nadjib Badache, Ad Hoc Networks Routing Protocols and Mobility, The International Arab Journal of Information technology, Vol 3, No. 2, April-2006

[3] Feng Li and Jie Wu, Mobility Reduces Uncertainty in MANETs, INFOCOM. 26th IEEE International Conference on Computer Communications. IEEE, 2007

[4] Bhavyesh Divecha, Ajith Abraham, Crina Grosan and Sugata Sanyal, Impact of Node Mobility on MANET Routing Protocols Models, Journal of Digital Information Management, Volume 5, Number 1, pp. 19-24, 2007

[5] Andrea E.F. Clementi, Francesco Pasquale, and Riccardo Silvestri, MANETS: High mobility can make up for low transmission power*, CoRR abs/0903.0520: (2009)

[6] Nils Aschenbruck, Matthias Frank, Peter Martini, Jens Tolle, Human Mobility in MANET Disaster Area Simulation- A Realistic Approach, Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks (LCN'04)

[7] R. Timo, K. Blackmore, and L. Hanlen, Strong Stochastic Stability for MANET Mobility Models, Networks. ICON. 15th IEEE International Conference on 2007.

[8] Xianren Wu, Hamid R. Sadjadpour and J.J. Garcia-Luna-Aceves, Link Lifetime as a Function of Node Mobility in MANETs with Restricted Mobility: Modeling and Applications, In proceeding of: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops, 2007.

[9] Yasser Kamal Hassan1, Mohamed Hashim Abd El-Aziz2, and Ahmed Safwat Abd El-Radi1, Performance Evaluation of Mobility Speed over MANET Routing Protocols, International Journal of Network Security, Vol.11, No.3, PP.128{138, Nov. 2010

[10] Baldev Ram Mali, N.C. Barwar, Effect of Mobility on Performance of MANET Routing Protocols under Different Traffic Patterns, IP Multimedia Communications. Published by Foundation of Computer Science, New York, USA, 2011

[11] Gang Lu, Gordon Manson and Demetrios Belis, Mobility Modeling in Mobile Ad Hoc Networks with Environment-Aware, Journal of networks, Vol. 1, NO. 1, May 2006

[12] Juan-Carlos Cano, carlos T. Calafate, Manuel P. Malumbres, and Pietro Manzoni, Evaluating the performance impact of group mobility in MANETs, Conference on Parallel, 15-17th September 2004, Almeria, Spain.

[13] Theus Hossmann, Mobility Prediction in MANETs, Master's Thesis - Winter Term 2005/2006

[14] Xiaofeng Lu, Yung-Chih Chen, Pietro Liò, Don Towsley, Modeling Mobility from Military MANET Traces, Proceedings of the Second Annual Conference of the International

Technology Alliance, London UK, September 2008

[15] N. Kumaresh, An Empirical Approach - Distributed Mobility Management for Target Tracking in MANETs, International Journal of Computer Applications (0975 – 8887) Volume 2 – No.6, June 2010

[16] Suprio Ray, Realistic Mobility for MANET Simulation, Thesis for Master of Science, Regional Engineering College, Trichy, 2003

[17] Illya Stepanov, Pedro José Marrón, and Kurt Rothermel, Mobility Modeling of Outdoor Scenarios for MANETs, Proceedings of the 38th Annual Simulation Symposium, 2005

[18] Maoning Wang, MANET Global Connectivity and Mobility Management Using HMIPv6 and OLSR, Thesis for Master of Computer Science, School of Computer Science Carleton University Ottawa, Ontario, 2003

◈◈◈

The Significance of Data Security in Cloud: A Survey on Challenges and Solutions on Data Security

Ashalatha R & Vaidehi M.

Computer Science, DSCE, Bangalore, India
RIIC, DSI, Bangalore, India
E-mail : ashalatha.dsce@gmail.com, dm.vaidehi@gmail.com

Abstract - Cloud Computing has changed the entire operations of IT industries mainly the security issues of essential data. The cloud computing has benefited the IT (Information Technology) industries with less infrastructure investment and maintenance costs. The cloud provides services like Infrastructure-as-service (IaaS), Platform-as-Service (PaaS) and Software-as-service (SaaS) to its clients with vast storage capacity by ensuring data security. Security is an essential service to be provided in public cloud and hybrid cloud environment where the data can be easily hacked or tampered with. The paper aims to provide a comprehensive review on the essentiality of Security- as- Service in cloud computing environment and also presents the importance of data security in cloud. The survey is done on the challenges and solutions provided on the cloud for the data security.

Keywords - *Cloud Computing, security, public cloud, hybrid cloud.*

I. INTRODUCTION

Cloud computing is an growing field in the IT (Information Technology) industries mainly for the security community because cloud architectures are literally popping up all over the world [1]. The security of cloud computing services is an issue which is delaying its adoption in today's technology world. Various issues barring the adoption of cloud computing are in large part to the private and public sectors which unease surrounding the external management of security-based services. It is the nature of cloud computing-based services, private or public sectors that promote the external management of provided services to its clients. This delivers the great incentive to cloud computing service providers for security reasons [2].

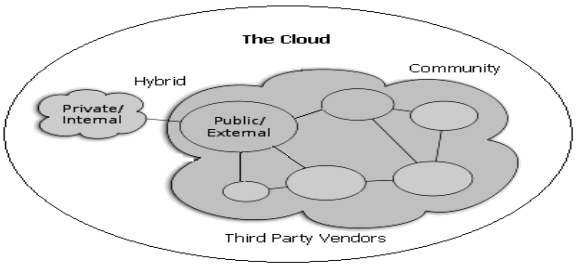


Fig 1. Types of cloud computing

The figure 1 depicts various types of cloud computing models which are Public, Private, Hybrid and Community cloud. The public cloud is based on the standard computing model where the resources like applications and storage devices are made available to the public through World Wide Web. These services may be free or on a pay basis. The community cloud shares the infrastructure around several organizations which can be managed and hosted internally or by third party providers. The Hybrid cloud is a combination of two or more clouds which are bound together to offer multiple deployment models. It can also be defined as multiple cloud systems that are connected in a way that allows programs and data to be moved easily from one deployment system to another. The Private cloud is infrastructure operated solely for a single organization whether managed internally or by third-party vendors.

A. The Cloud Models or Layers:

Software as a Service (SaaS) provides a complete application of various software's which is offered to the customer over Internet. A single instance of the service runs on the cloud & multiple end users are serviced here. On the customer's side, there is no need for upfront investment in servers or software licenses, while for the provider, the costs are lowered, since only a single application needs to be hosted & maintained [3].

Platform as a Service (PaaS): provides a layer of software platform or the development environment which is encapsulated & offered as a service. The layer provides the storage of various applications and its

deployment. The customer has the freedom to build his own applications, which run on the provider's infrastructure. To meet manageability and scalability requirements of the applications, PaaS providers offer a predefined combination of OS and application servers, such as LAMP platform (Linux, Apache, MySql and PHP) software's.

Infrastructure as a Service (IaaS): provides basic storage and computer infrastructure as standardized services over the network. Servers, storage systems, networking equipment, data centre space etc. are pooled and made available to handle workloads. The customer would typically deploy his own software on the infrastructure layer without the need of the servers. For the IaaS, we need to pay only when we are using the service and thereby saving the infrastructure cost.

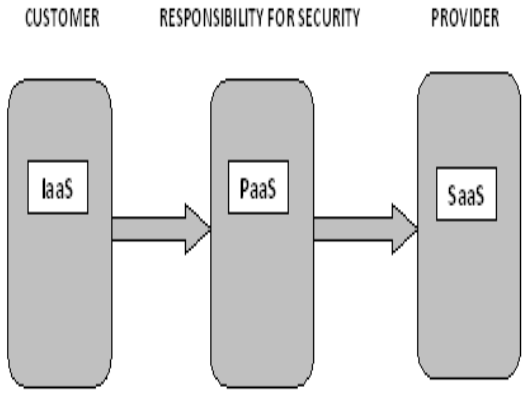


Fig 2: Different Cloud Models

Figure 2 depicts the three cloud computing models which include Infrastructure, platform and software services. These services are provided by the cloud providers for the clients who rely on cloud for storage and other types of services.

B. Protection of Cloud:

A Secure cloud is always a reliable source of information for protecting the cloud which is a very important task for security professionals who are in charge of the cloud. Some of the ways by which a cloud can be protected are Protection of data, making sure data is available for the customers, delivering high performance for the customers, using Intrusion Detection System on cloud to monitor any malicious activities. To make sure the application used by the customer is safe to use, the providers must provide a support system for the customer and the customers should be able to recover any loss of data in the cloud. There should be a good degree of encryption given by the providers to the customers wherein only the customer should be able to access the data and not the malicious users [4].

C. Organization: The rest of the paper is organized as follows: Section II reviews the related work, Section III gives the security issues associated with cloud computing. Section IV gives the conclusion and future work.

II. RELATED WORK

Mihaela Ion et.al [5] says the cloud computing gives huge amount of data storage at very low cost which uses the virtualization techniques. The most confidential data are protected using encryption on the cloud. This scheme allows complex encrypted queries for encrypted data in a multi programming environment. It uses SQL queries on encrypted databases in multiuser program. It uses complex queries which encrypts table name and attribute names. It has access control which is a Key-Policy Attribute Bases Encryption (KP-ABE) is used for encryption.

Ryan K L Ko et.al [6] discusses the key issues and challenges to achieve a trusted cloud by its service users by making use of detective controls and shows the framework called as TrustCloud which gives the accountability in cloud computing through technical and policy based services. The various components used in cloud computing here are Security, Privacy, Accountability and Auditability.

Muhammad Rizwan Asghar et.al [7] addresses the issues like enforcing security policies in untrusted environment by protecting the policy confidentiality. ESPOON (Encrypted Security Policies for OutsOurced eNvironments) provides separation between security policies and the enforcement mechanism. The approach here uses an encryption scheme to protect confidentiality of the policies which uses policy based method. This approach has two main phases in the life cycle which is policy deployment phase and the policy evaluation scheme. The algorithm used here is based on multi-user searchable data (SDE) scheme which is used in each phases of policy life cycles.

Danwei Chen et.al [8] proposes a secure data strategy in cloud computing. The paper is based on fundamental theories of k equations in algebra, n congruence surplus principle in elementary number theory, and the Abhishek's online data storage algorithm. The strategy splits data d into k sections using the data splitting algorithm. It also ensures high data security by simplifying k equation solutions and guarantees highly reliable data using the coefficients which are generated by this algorithm.

III. SECURITY ISSUES ASSOCIATED WITH CLOUD COMPUTING

1. Privacy issues: Privacy means the accountability of organizations to data subjects, as well as the transparency to an organization's practice around personal information [9]. In privacy issue, the provider must make sure that the customer's personal information is very well secured from other operators. Due to the external servers' attachment, the provider should make sure who is accessing the data and who is maintaining the server and thereby enabling the provider to protect the customer's personal information. These providers ensure that all essential data or the sensitive information are masked and only the authorized clients have access to data in its entirety. The digital identities and credentials must be protected because of the data that the provider collects or produces about customer activity in the cloud. Cloud stores the data from many clients and can run the data mining algorithm to get large amount of information on clients [10]. Privacy rights are related to collection, use, disclosure, storage and destruction of personal data.

2. Control issues: Controlling the data from third party brokers is an important issue for all outsourced networked applications and storage, but cloud architectures have some specific issues which are distinct. Physical control of the private cloud equipment is more secure than having the equipment off site and under someone else's control. The ability to visually view the data links and access ports is essential to ensure the data links which are not be compromised. Allowing a third-party service to take custody of personal documents raises awkward questions about control and ownership [11]. The issues of privacy and control can only be assured with tight service-level agreements (SLAs).

3. Physical access: In the physical access issue, the main consideration of this issue is the organization's staffs who are not having physical access to the machines storing and processing the data. The issue of unknown third parties who are having physical access to the machines is also an important issue. One simple solution is to use in-house "private clouds" [12]. The value of filtering a packet-sniffer output to specific services as an effective way to address security issues shown by anomalous packets directed to specific ports or services [13]. Another solution for accessibility vulnerabilities is to shut down unused services, keep patches updated and reduce the permissions and access rights of applications and users.

4. Security: Security is an essential component of strong privacy safeguards in all online computing environments. Cloud customers and business providers both are willing to use online computing only if they trust that, their data will remain private and secure. Addressing of privacy challenges in the evolving cloud computing is a very difficult task from many a years. The extensive experience has helped to develop well-defined business practices, privacy policies, and security measures that govern the cloud computing ecosystem. Recognizing that the cloud poses some new security and privacy challenges, the current policies and practices provide a solid foundation for addressing privacy issues and enabling greater trust in the Internet [14].

Types of Security levels are listed below.

a) Vendor level Security: The Vendor should make sure that the server is well secured from all the external threats they come across. A Cloud is good only when there is a good security provided by the vendor to the customers.

b) User level Security: Even though the vendor has provided a good security layer for the customer, the customer should make sure that because of its own action, there shouldn't be any loss of data or tampering of data for other users who are using the same cloud.

c) Physical and personnel security: Providers ensure that the physical machines are adequately secure and that access to these machines as well as all relevant customer data is not only restricted but that access is documented.

d) Security of data at data center: Organizations are skeptical about the data security because of "third party vendor and multi tenancy". Choice of cryptographic and hash algorithms used, how it works at transport layer and how data protected from other tenants being the center issue. *Multi-tenancy* is the obvious choice for the cloud vendors for scalability but large enterprises see it as a weapon to exploit their huge database.

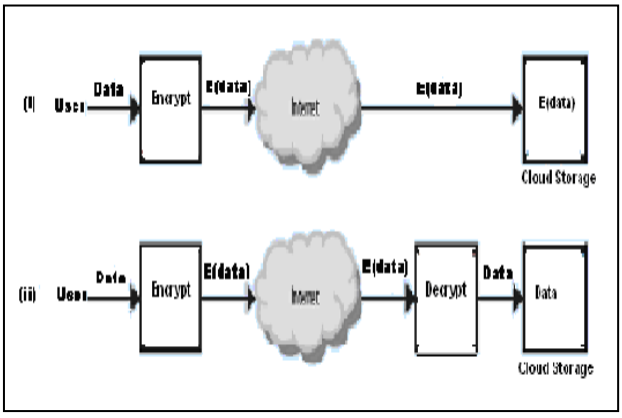


Fig 3: Encryption techniques

Figure 3 depicts two different encryption scenarios existing for cloud computing. In the first part, the data remains decrypted at the cloud storage site thereby

preventing the unauthorized access through the Internet. The cloud provider also cannot access the original data. In the second part of the figure 3, the data are decrypted by the cloud provider to enable necessary operations on the data.

Security Solutions: There are several groups interested in developing standards and security for clouds and cloud security. The Cloud Security Alliance (CSA) is gathering solution providers, non-profits and individuals to enter into discussion about the current and future best practices for information assurance in the cloud. The Cloud Standards website is collecting and coordinating information about cloud-related standards under development by other groups [15]. The Open Web Application Security Project (OWASP) maintains a "top 10" list of vulnerabilities to cloud-based or Software as a Service deployment models which is updated as the threat landscape changes [16]. The Open Grid Forum publishes documents containing security and infrastructural specifications and information for grid computing developers and researchers [17].

IV. CONCLUSION

The largest gaps between cloud-security practice and cloud-security research lies in the fact that the assumptions in the research leave out some very important differences between cloud security and virtual machine security. One of the pieces of the framework might be developing a way to monitor the cloud's management software, and another might be development of isolated processing for specific clients' applications. Having a way to tell whether the virtual machines in the cloud are patched properly would also be a useful part of the framework. People's behavior can be tracked and monitored; for instance whether people allow the automated patching software to run, or updating anti-virus software definitions (on virtual machines running operating systems that are susceptible to viruses, worms and other such malware), or whether people understand how to harden their virtual machines in the cloud. Most of the companies have been addressing many of these issues since 1994, when they delivered their first online services for consumers and enterprises. The breadth of experience has shaped the company's privacy principles, corporate privacy policy, product and service development, and overall business practices. These components anchor commitment to maintaining the highest standards of privacy and security in our online services and to partnering with other industry leaders, governments, and consumer organizations to develop globally consistent privacy frameworks that enable the expansion of the economic and social value of cloud-based computing.

V. FUTURE WORK

The Security optimizations can be enhanced to shared accreditation, validation of customer applications, integrating software as a service, accessing federated and shared services, varying interpretations of security guidelines [18].

REFERENCES

[1] Wolf Halton, "Security Solutions for Cloud Computing", July 15, 2010.

[2] Wolf Halton, Opensource and security on "Security Issues and Solutions in Cloud Computing", July 25, 2010, wolf in cloud computing, Tech Security.

[3] Ponemon Institute and CA "Security of cloud computing Users: A study of Practitioners in the US & Europe". May 12, 2010.

[4] Lord CrusAd3r "Problems faced by Cloud Computing".

[5] Mihaela Ion, Giovanni Russello, Bruno Crispo, "Enforcing multi-user access policies to encrypted cloud databases" 2011 IEEE International Symposium on Policies for Distributed Systems and Networks.

[6] Ryan K L Ko, Peter Jagadpramana, Miranda Mowbray, Siani Pearson, Markus Kirchberg , Qianhui Liang , Bu Sung Lee, "TrustCloud: A Framework for Accountability and Trust in Cloud Computing" 2011 IEEE World Congress on Services.

[7] Muhammad Rizwan Asghar, Mihaela Ion, Bruno Crispo, "ESPOON Enforcing Encrypted Security Policies in Outsourced Environment", 2011 Sixth International Conference on Availability, Reliability and Security.

[8] Danwei Chen, Yanjun He, "A Study on Secure Data Storage Strategy in Cloud Computing", Journal of Convergence Information Technology, Volume 5, Number 7, September 2010.

[9] Privacy in the Cloud Computing Era: A Microsoft Perspective - November 2009.

[10] Bharat Bhargava, Anya Kim, YounSun Cho, "Research in Cloud Security and Privacy".

[11] Hayes, B. (2008). Cloud computing. Commun. ACM, 51(7), 9-11. Retrieved from http://portal.acm.org.library.capella.edu/ft_gateway.cfm? id=1364786&type=html& coll=portal& dl=ACM&CFID= 0867670& CFTOKEN=24312614.

[12] Milne, J. (2010, February 9). Private cloud projects dwarf public initiatives.Retrievedfrom http://www.cbronline.com/news/private_cloud_projects_dwarf_public_initiatives_281009: http://wolfhalton.info/2010/06/25/security-issues-and-solutions-in-cloud-computing/ #ixzz1ruqU5eIM

[13] KrÃ¼gel, C., Toth, T., & Kirda, E. (2002). Service specific anomaly detection for network intrusion detection. In Proceedings of the 2002 ACM symposium on Applied computing (pp. 201-208). Madrid, Spain: ACM. Retrievedfrom:http://portal.acm.org.library.capella.edu/citation.cfm?id=508835&dl=GUIDE&coll=GUIDE&CFID=80867670&CFTOKEN=24312614.

[14] Eystein Mathisen, "Security Challenges and Solutions in Cloud Computing", 5th IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST 2011), 31 May -3 June 2011, Daejeon, Korea

[15] cloud-standards.org/ Clouds Standards, 2010. Weblink: http://cloud standards.org/wiki/index.php?title=Main_Page

[16] OWASP,2010Weblink: https://www.owasp.org/index.php/Top_10_2010

[17] Open Grid Forum, 2010. Weblink: www.ogf.org/

[18] Henry J. Sienkiewicz, "Cloud Computing: A perspective" Defense Information Systems Agency - April 2009.

◈◈◈

Cylindrical-Sector Based Location Management In Mobile Ad Hoc Network

Ramya.A.R, R.Anand & Jitendranath Mungare

Dept. of Computer Science & Engineering, CMR Institute of Technology, Bangalore, India
E-mail : haiiamramya@gmail.com

Abstract **-** The proposed system exhibits a novel scheme of location based services based on spherical shape data aggregation techniques in mobile adhoc network. The system basically highlights the issue of design for scalable routing protocols. Location based routing services is the prime focus of the design of the experimental test bed. The adopted technique assumes that mobile nodes moves around and relay each other data packet in a spherical scenario, which is further classified into sector and cylinders. A geographic forwarding technique is considered for this purpose. Location server updates the packets while location query is conducted by forwarding a query packet in dual oposite directions along the sectors. The efficiency of the protocol is maintained by considering scalability, mobility, delivery ratio, different network size, as well as byte overhead.

Keywords **-** *component; Mobile adhoc network, location based services, geographic routing.*

I. INTRODUCTION

Mobile Adhoc Networks [1] [2] [3] are collective arrangement of mobile nodes that can communicate with one another without the aid of any centralized point. Adhoc networks make practical and effective use of multihop radio relaying and radio communication channel. It [4] is very important for one mobile host to enlist the aid of other hosts in forwarding a packet to its destination, due to the limited range of each mobile node's transmissions. With the enhancement of technology, this network could be managed by end users rather than single authority and they may be used for extremely sensitive applications. In adhoc networks, node mobility is an important issue due to adhoc characteristics such as dynamic network topology, shared medium, limited bandwidth, multihop nature and security etc. Thus, there is requirement of effective mobility management scheme i.e. seamless mobility in adhoc networks. Seamless mobility provides easy access and effective communication among nodes present in the network. A Mobile Ad Hoc network (MANET) is an autonomous system of mobile nodes connected by wireless links. In a MANET it is assumed that the nodes are free to move and are able to communicate with each other, often through multi-hop links, without the help of a fixed network infrastructure. The network topology is dynamic. The movement of a node out of or into the Communication range of other nodes changes not only its neighbour relationships with those other nodes, but also changes all routes based on those relationships. Signaling overhead traffic for establishing and maintaining routes in a MANET is proportional to the rate of such link changes. Thus the performance of a MANET is closely related to the efficiency of the routing protocol in adapting to changes in the network topology and the link status [5], [6] For the performance evaluation of a routing protocol for a MANET, it is imperative to use an appropriate mobility model to simulate the motion of the nodes in a network [7]. In this paper we present some mobility models that have been proposed, or used in, the performance evaluation of Ad Hoc network protocols. The models presented are the random waypoint mobility model [5], the random Gauss-Markov model [8], [9], and the reference point group mobility model [10].

Mobility models in adhoc networks depict [10] movement pattern of mobile users and how their location, velocity, speed, direction and acceleration change over time. In these networks, mobile nodes communicate directly with each other. Communication between two nodes does not produce effective results if both nodes are not in same transmission range. This problem can be resolve by using intermediate nodes with routing. Thus, routing is very important in mobile adhoc networks where mobility models must be evaluated with respect to end to end delay and efficient data transmission. Mobility models are intended to focus on individual movement patterns due to point to point communication in cellular networks [11-12] [9] whereas adhoc networks are designed for group communication.

Such models [13] are suggested to maintain movement, and efficient transmission among nodes in real life applications. In addition to this, these models are mainly focused on the individual motion behavior between mobility era with minimum simulation time in which a mobile node moves with constant speed and direction. These models represent the features of the mobile nodes in an adhoc network like speed, direction, distance and node movement. Mobility models [7] can be categorized based on the following criteria which is based on dimension, scale of mobility, randomness, geographical constraints, destination oriented and by changing parameters (discussed in next section). Generally, there are two types of mobility models (i) Trace based mobility models and (ii) Synthetic mobility models. Trace models provide mobility patterns based on deterministic approach whereas synthetic models presents movements of mobile nodes in realistic manner.

In section 2 we give an overview of related work which identifies all the major research work being done in this area. Section 3 highlights about the proposed system. Module description is discussed in Section 4 followed by implementation and analysis in Section 5. Section 6 discusses some concluding remarks.

II. RELATED WORK

P. Krishna et.al [14] present Location management is an important problem in distributed mobile computing. Location management consists of location updates, searches and search-updates. An update occurs when a mobile host changes location. A search occurs when a mobile host needs to be located. Upkar Varshney [15] presented the requirements of several location-intensive mobile commerce applications and integrated location management architecture to support these requirements. They also showed how the proposed location architecture could be used to support future m-commerce applications. Kuo-Hsing Chiang and Nirmala Shenoy [16] present in this work, a novel two-dimensional (2-D) random-walk mobility model is proposed, which can be used for studying and analyzing the location-area crossing rate and dwell time of mobile users in wireless networks.

Jadwiga Indulska et.al [17] presents a location management system able to gather process and manage location information from a variety of physical and virtual location sensors. Travis Keshav [18] included in this survey indicate, static LM schemes are becoming increasingly out of date. Dharmesh Niranjan and A. K. Vatsa [19] proposed a novel architecture and mechanism for Context Based Location management (CBLM) using mobile agent through sensors processed parameters values that determined the cluster head creation and updates if required and Location Management. Wilde, Erik, and UC Berkeley [20] present a concept for representing location vocabularies, matching and mapping them, how these vocabularies can be used to support better privacy for users of location-based services, and better location sharing between users and services.

James Cowling [21] presents a complete and novel location management scheme, addressing weaknesses present in previous proposals, while maintaining a high level of implementation feasibility. Kyandoghere Kyamakya and Klaus Jobmann [22] actualizes the classification of location management methods published up to now and presents results of a related extensive performance comparison of the most important paradigms for location management in cellular networks.

Jingyuan Zhang [23] present in a cellular network, a service coverage area is divided into smaller areas of hexagonal shape, referred to as cells. The cellular concept was introduced to reuse the radio frequency.

Jingyuan Zhang [24] identified these weak points and discussed the schemes to eliminate or reduced these weak points. This chapter has reviewed several individualized and dynamic schemes that are able to more evenly distribute the signaling burden of location updates.

Bhavneet Sidhu and Hardeep Singh [25] surveyed research on location management in cellular networks. John Sucec and Ivan Marsic [26] proposed that by applying the distributed database selection technique of GLS, a hierarchical location management scheme may be realized for MANETs based on link state routing that equitably distributes location server functionality among network nodes. Mujtaba Khambatti and Sarath Akkineni [27] describe two novel location management strategies: hierarchical and de-centralized.. Sumesh J. Philip et.al [28] uses a theoretic framework to show the asymptotic scalability of three location management protocols. They also carry out extensive simulations to study the performance of these protocols under practical considerations. Demin LI [29] presents an approach to the total cost estimation of variable velocity mobile location management for ad hoc mobile networks with missing measurements.

III. PROPOSED SYSTEM

The main aim of the proposed system is to design an architectural framework for modeling a robust scalable location management service for routing in mobile adhoc network which ensures very low overhead and secure process of routing. In this project work, the focus is on presenting the proposed technique, a scalable

location management service for routing in mobile ad-hoc networks that uses a circular system model and also provide security for the same. It has to show that the proposed system saves the communication overhead when all nodes are constantly up without losing the performances compared with the previous efficient location-based routing protocols especially in dense networks. The functional requirements of the proposed system are as follows:

- The system should able to deploy the nodes in random manner in the 2 cylinders as well as total of 24 sectors in the cylinders.
- The system should able to design a mobility model using random waypoint model.
- The system should able to implement the routing protocols for location update packages that moves along the cylinder and location query packages move along the sector.
- The system should have an efficient updating process for the location server and packets of information being transacted by them.
- The system should not increase the network overhead (byte overhead) if the proposed algorithms are implemented in real sense.

This method supposes that nodes moves around and relay each other packet in a circular environment. This circle is divided into equal sectors and cylinders. Each area has a unique id. The proposed system architecture of the project work is as shown below:

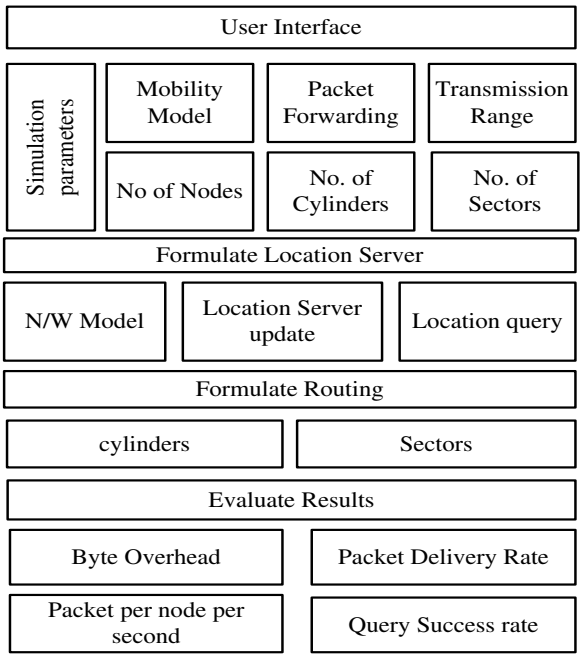


Fig. 1 : Proposed Architecture

The proposed system assumes a central point in each area. Nearest nodes to this central points in each area are selected as local server. Each local server is responsible for response to location queries for all owner member nodes. Location update packages move along the cylinder and location query packages move along the sector. These update and query propagation method decrease system response time, because the query packet traversed only one sector. One of the sectors is considered as basic sector. Location servers in basic sector periodically start to send location update packets in two directions. These packages gather updates of local servers from each area. All packages are sent by geographic method.

A new node has to be initially authenticated by each of its neighbors to join the network. Once that has been accomplished, each packet sent by the node to its one-hop neighbor is authenticated by the neighbor using a packet authentication tag. The one-hop neighbor then replaces the tag with its own authentication tag and forwards the packet to its neighbor. This next neighbor verifies the new authentication tag as coming from its immediate neighbor and the process is repeated iteratively until the packet reaches its destination. Therefore, each packet is authenticated at every hop. This scheme has the advantage that is resistant to Denial of Service (DoS) attacks and sessions hijacking attacks such as man-in-the-middle attack.

An effective IDS is a key component in securing MANETs. Two different methodologies of intrusion detection are commonly used: anomaly intrusion detection and misuse intrusion detection. Anomaly-detection systems are usually slow and inefficient and are prone to miss insider attacks. Misused detection systems cannot detect new types of attack. Hybrid systems using both techniques are often deployed in order to minimize these shortcomings

IV. MODULE DESCRIPTION

The proposed system consists of following 3 modules:

A. System Model:

Network area is modeled as a circle-shaped. This circle is divided into cylinder and sector. Each area is created by intersection between sectors and cylinders. Each node at any moment can stay in one of these areas.

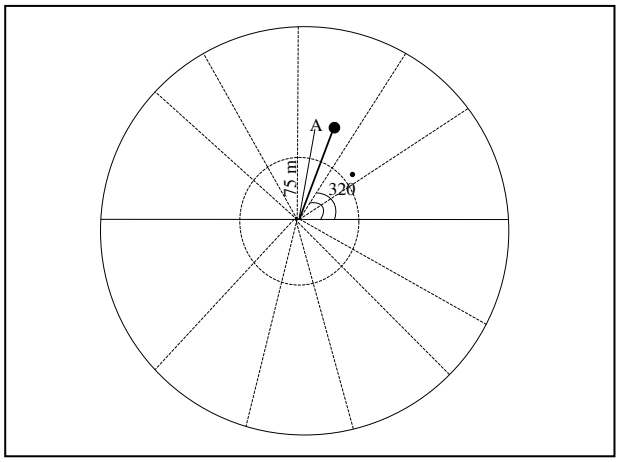


Fig. 2 : Network model in proposed system

Size of each zone must be set so that all nodes within that region will be in radio broadcast of local area server. Due to the local servers are located approximately at the center of each area. At the worst case the local server may not be located in the middle of area and may be in one of the corners. Then all the nodes within the area are in location server radio broadcasts. Each node with beaconing process can obtain direct neighbors locations. It calculates its own distance to area center and other neighbor nodes too. Closest node to area center point well is that location server. Therefore the local server of each area will be selected.

B. Location Server Update:

In the proposed work, servers are located in central areas. One sector considered as basic sector to start location server update process. Selection of basic sector is predefined in network model. Server update packet is sent periodically. These periods grows as the cylinder size grows. These update packets use synchronize aggregated method and we use this method with a little change. Location servers in each area in basic sector send two packets in clockwise and anti clockwise directions along on cylinder are on that. Each node that receive this packets if it is not server just forward packets to server on its area. Then apply member nodes location information changes and forward those to the next area. Update packets forward continues until packets reach to the basic sector again. Unlike conventional methods, so that once they attempted to gather information and then information collected will be sent. Receiving these update packets by server nodes in basic sector means that all location servers in cylinders are update and can respond to location queries.

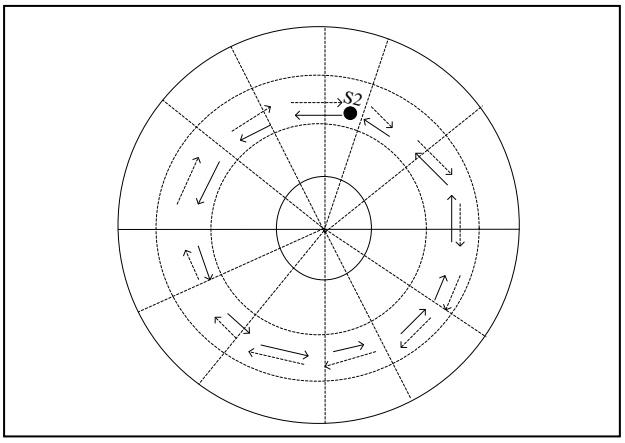


Fig. 3 : Servers update process in proposed system

C. Location Query:

Location query process in the proposed system is done by sending a query packet in two different directions along the sector. When a source node wants to send the data to a destination node, if it doesn't become aware of destination's location, it should send a query packet to its location server. If the local server doesn't have destination position, had to find that. Intermediate nodes that receive the query packet, if they themselves have the destination's location, reply to query and send destination's location to source node, otherwise resend query packet to previous direction. The destination node is located in one of the cylinder; therefore one of the local servers has information of destination location. So finally one of the two packages can be answered and the source node will be aware destination location.

To prevent the exit of location servers from the sector the first cylinder (cylinder number zero) and the last cylinder when receive query packet, don't resend it and if it cannot respond to query packet, it sends back an error message to requested node.

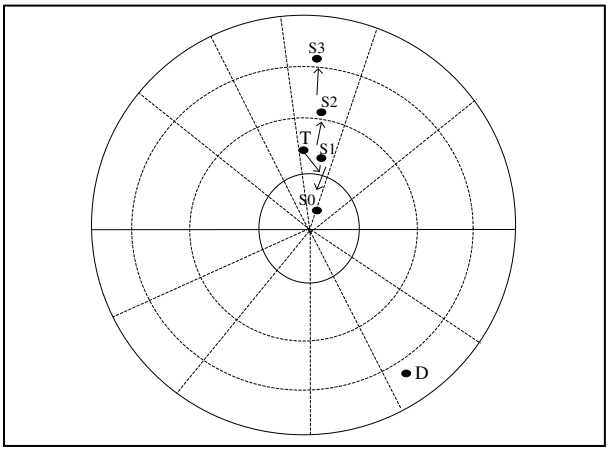


Fig. 4 : Location query process in proposed system

D. Security mechanism:

Two nodes authenticate each other using signed unforgeable certificates issued by a virtual trusted certificate authority. Multiple nodes will function collectively as a certificate authority. Authority and functionality of an authentication server is distributed across k nodes that collaboratively serve and provide authentication services. A central certificate authority is not reasonable in MANET architectures because they are vulnerable as a single-point of compromise and/or failure. A local trust model is necessary in order to perform any type of authentication services. A user is considered trusted if any *k* trusted entities claim so within a fixed time period.

These *k* entities are usually among the user's one-hop neighbors. K may be either a globally set parameter or may be a function of location (i.e., majority of each node's neighboring nodes). All authenticated nodes carry a certificate signed with the network secret key (SK). Authenticated nodes help each other forward and route packets, while unauthenticated nodes are denied access to network resources. Authenticated nodes also perform network monitoring for suspicious activity from their neighbors.

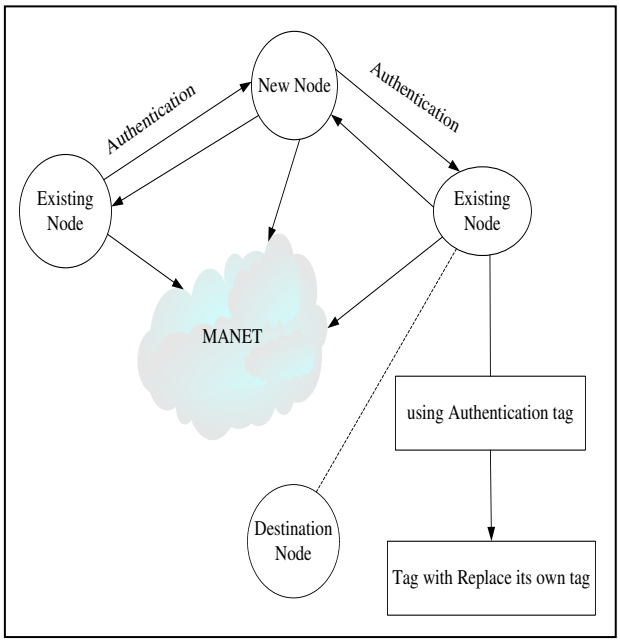


Fig. 5 : Hop by hop authentication mechanism

V. IMPLEMENTATION & ANALYSIS

The proposed system is implemented in 32 bit Windows OS with 1.84 GHz Processor. The design environment is selected in Matlab. The process flow chart of the implementation phase is as shown in Figure 6 and 7. The mobile nodes use IEEE 802.11 radio and MAC model with bit rate 11Mbps and each radio range is approximately a disc with a 250 meter radius. The size of each simulation universe is chosen to maintain an average node density of around 100 nodes per square kilometer as shown in figure 8. Each simulation runs for 300 simulated seconds. Each simulation runs for 300 simulated seconds. Each node moves using a random waypoint model. The node chooses a random destination and moves toward it with a constant speed chosen uniformly between one and a maximum speed (10 m/s unless noted otherwise). When the node reaches the destination, it chooses a new destination and begins moving toward it immediately. All simulations use a pause time of 0 second. For two location service protocols, we chose a beacon period of 2 seconds and the value update period was fixed at 25s.

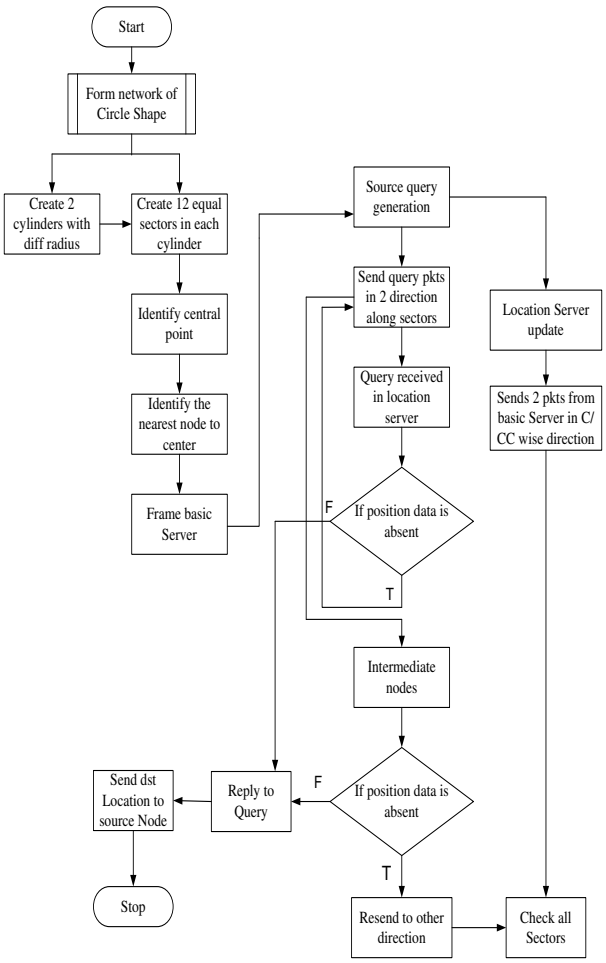


Fig. 6 : Implementation process flowchart.

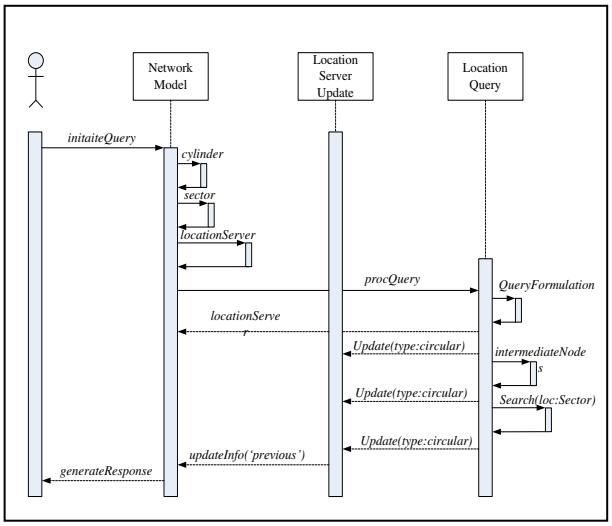


Fig.7 : Sequence Diagram

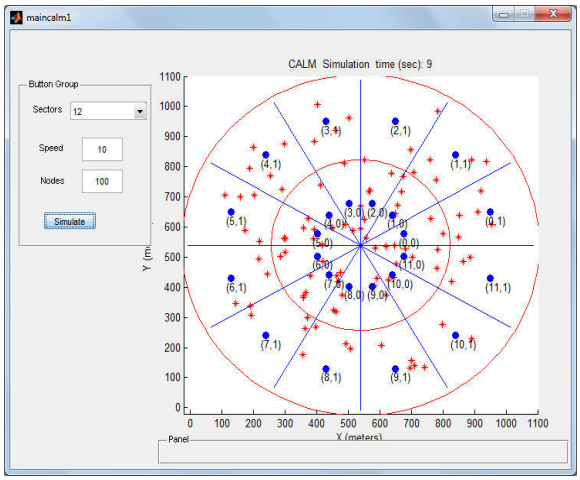


Fig. 8 : User interface

The metrics Location Service protocol overhead, Packet delivery ratio and Query success ratio for location services are the parameters considered for performance analysis. The protocol overhead includes the number of location service protocol packets or bytes that transmitted during the simulation time, note that this excludes the overhead due to beaconing. The query pattern is chosen to study the efficiency of the query and update mechanism of two protocols. Every node in the network initiates certain number of queries to look up the location of randomly destinations at times randomly distributed between certain intervals of time. Also, if a query is not successful, no retry is initiated. To measure the accuracy of the query reply, when the query reply is received, each source sends a single data packet of size in bytes to that destination using the replied location, this for routing done after finding a route to destination. Hence the performance is highly optimized for the proposed system

VI. CONCLUSION

We have presented a novel technical where a scalable location management service for routing in mobile ad-hoc networks that uses a circular system model is shown. We have shown that the proposed system saved the communication overhead when all nodes are constantly up without losing the performances compared with the previous efficient location-based routing protocols especially in dense networks. This is achieved thanks to the use of "Synchronized Aggregation" technique.

REFERENCES

[1] C.E. Perkins, " Ad hoc Networking", Addison Wesley 2001.

[2] C.K. Toh, " Adhoc Mobile Wireless Networks: Protocols and Systems", Printice Hall, New Jersy, 2002.

[3] Dr. Yogesh Chaba and Naresh Kumar Medishetti, " Routing Protocols in Mobile Ad hoc Network – A Simulation Study Final", JCS Vol 1. No. 1. August 2005.

[4] David B. Johnson, David A. Maltz, " Dynamic Source Routing in Ad hoc Networks", In Mobile computing, Vol. 353 (1996), pp. 153-181.

[5] J. Broch, D.A. Maltz, D.B. Johanson, Y.C. Hu, and J. Tetcheva, " A Performance comparision of multi-hop wireless ad hoc network routing protocols", IEEE/ACM, Mobicom'98, pp.16-28, 1998.

[6] C.E. Perkind, E. M. Royer, s.R. Das, and M.K. Marina, " Performance comparison of two on-demand routing protocols for ad hoc networks", IEEE Pers. Commun. Vol.8. No.1. pp. 16 -28, 2001,

[7] Byung Jae Kwak, Nah-Oak Song, and Leonard E. Miller, "A standard measure of mobility for evaluating mobile ad hoc network performance", IECE TRANS. COMMUN. E86-B, 2003.

[8] D. Shukla, Mobility models in ad hoc networks, Master's thesis, KReSIT-ITT Bombay, Nov. 2001.

[9] T. Camp, J. Boleng, and V. Dvies, " A survey of mobility for ad hoc networks research, Wireless Communication and Mobile Computing (WCMC); Special issue on Mobile Ad hoc Networking: Research, Trends and Applications, pages pp. 483-502, 2002.

[10] X. Gong, M. Gerla, G. Pei, and C. C. Chiang." A group moibility model for an ad hoc wireless

network", Proc. ACM/IEEE, MSWiM'99, pp. 53-60, 2000.

[11] Xiaoyan hong, Mario Gerla, Guangya Pei, and Ching-Chuan Chiang, " A gourp mobility model for ad hoc wireless networks", August 1999 in Proceedings for the ACM International workshop on Modeling and Simulation of Wireless and Mobile systems (MSWim)

[12] D. Johnson and D. Tse, " the dynamic source routing protocol for mobile ad hoc networks (DSR)", Feb. 2002, IETF internet Draft draft-ietf-manet-dsr-07.txt.

[13] F.P. Kelly, A.K. Maulloo and D.K.H. Tan. " Rate control for communication networks shadow prices, proportional fairness and stability", Journal of the Operational Research society, 49(3), 237-252, March 1998.

[14] P. Krishna, Nitin H. Vaidya, Dhiraj K. Pradhan "Location management in distributed mobile environments" Date of Conference: 28-30 Sep 1999, P-81 – 88

[15] Upkar Varshney, Location Management for Mobile Commerce Applications in Wireless Internet Environment Georgia State University ACM Transactions on Internet Technology, Vol. 3, No. 3, August 2003, Pages 236–255.

[16] Kuo-Hsing Chiang and Nirmala Shenoy, Associate Member, IEEE A 2-D Random-Walk Mobility Model for Location-Management Studies in Wireless Networks IEEE Transactions on Vehicular Technology, Vol. 53, No. 2, March 2004

[17] Jadwiga Indulska, Ted McFadden, Matthias Kind, Karen Henricksen "Scalable Location Management for Context-Aware Systems"

[18] Travis Keshav "Location Management in Wireless Cellular Networks". Retrieved from: http://www.cse.wustl.edu/~jain/cse574-06/ftp/cellular_location.pdf

[19] Dharmesh Niranjan, A. K. Vatsa Context Based Location Management for MANET in Disaster Area using Mobile Agent IJCST Vol. 2, Issue 4, Oct . - Dec. 2011

[20] Wilde, Erik, "UC Berkeley Location Management for Mobile Devices" Publication Date: 02-01-2008 Publication Info: Recent Work, School of Information, UC Berkeley

[21] James Cowling "Dynamic Location Management in Heterogeneous Cellular Networks" October 2004 Advanced Networks Research Group School of Information Technologies University of Sydney Australia

[22] Kyandoghere Kyamakya and Klaus Jobmann "Location Management in Cellular Networks: Classification of the Most Important Paradigms, Realistic Simulation Framework, and Relative Performance Analysis" IEEE Transactions on Vehicular Technology, Vol. 54, No. 2, March 2005.

[23] Jingyuan Zhang, "Location Management in Cellular Networks," Handbook of wireless networks and mobile computing, John Wiley & Sons, Inc. New York, NY, USA, ISBN:0-471-41902-8, 2002

[24] Jingyuan Zhang, "Location Management in Personal Communication Systems" Computer Science Department, University of Alabama Ivan Stojmenovic, EECE, University of Birmingham, UK SITE, University of Ottawa, Ottawa, Ontário, Canada.

[25] Bhavneet Sidhu, and Hardeep Singh "Location Management in Cellular Networks" World Academy of Science, Engineering and Technology 25 2007

[26] John Sucec and Ivan Marsic "Location Management for Hierarchically Organized Mobile Ad hoc Networks" Department of Electrical and Computer Engineering Rutgers University,

[27] Mujtaba Khambatti and Sarath Akkineni "Location Management In Peer-To-Peer Mobile Adhoc Networks" Computer Science and Engineering Department Arizona State University Tempe, AZ 85287-5406 Technical Report May 2002

[28] Sumesh J. Philip, Joy Ghosh, Swapnil Khedekar, and Chunming Qiao "Scalability Analysis of Location Management Protocols for Mobile Ad hoc Networks" Department of Computer Science and Engineering State University of New York at Buffalo, Amherst, New York 14260

[29] Demin LI, Jiacun Wang, Jie Zhou, and Zidong Wang "Location Management Cost Estimation for Ad Hoc Mobile Networks with Missing Measurements" International journal of intelligent control and systems Vol. 11, No. 1, March 2006, 11-16

◈◈◈

A Software Approach for showing the performance of Cognitive Radio based Wireless Mesh Networks

K. Sandeep & K. Kalyan Krishnaiah

Department of Computer Science & Engineering, SITE, Tadepalligudem, AP, India.
E-mail : sandeepkone@gmail.com, kalyan97cse@gmail.com

Abstract - An application that presents new approach to wireless mesh networks spectrum utilization and management by Cognitive Radio along with the simulation results in java environment. In this work, the potential benefits in terms of QoS provided to users and the efficiency of resource utilization are quantified in a system consisting of a collection of one or more service providers. The objective of our simulation study done in java language is to shown that the Cognitive radio abilities provide an advantage over the classical network, either by improving QoS through increasing the probability of accepting correction requests, or by reducing the resources needed to fulfill the QoS requirements of users. This advantage is gained without impacting the service of primary clients. We also show that virtual wireless networks can be created to support volumes of users.

Keywords - *WMN, Specturm sensing, Cognitive Radio, wireless, QoS.*

I. INTRODUCTION

Wireless Mesh Networks(WMN) is an attractive technology which have been deployed, delivering services to cities, rural areas and organizations [1], [2], [3], [4]. As most WMN nodes do not require a wired connection to the internet, nodes can be installed quickly and in expensively on their placement with fewer conditions. They instead depend on wireless communication to interconnect individual access points, forwarding traffic over multiple hops to its destination. The wireless resources in WMN are scare. The WMN's multihop traffic must be transmitted over a limited amount of available spectrum. When combined with the openly shared nature of the medium, these factors limit the traffic capacity a WMN can support. It is, therefore, imperative that WMNs effectively utilize all the wireless resources available to them.

A new approach to managing wireless resources is Cognitive Radio (CR). Although the complete CR picture includes a wide-ranging vision of spectrum management combining perception, coordination, and intelligence, initial works are more limited, focusing on capitalizing on wasted spectrum. Under the current static spectrum allocation model, wireless channels are allocated to specific purposes or users. Primary users of a channel have exclusive rights to that channel. Even if the channel is idle and resources are being wasted, other wireless users are legally prevented from using the channel. In view of this, despite the demand for and expense of the resource, overall utilization is actually quite low. While some channels in some locations are heavily used, even congested, there are very often others that are essentially unused. The current system for resource allocation creates, therefore, many holes of unused spectrum due to static channel assignment, which prevents the spectrum bands from being used by anybody other than the primary user.

These spectrum holes are potential opportunities for making smarter use of wireless resources. Regulatory bodies have recognized this opportunity and acknowledged the need for a new system for spectrum management, one that is more flexible, dynamic, and efficient [5]. This impetus, combined with advances in radio technology and software-defined radio, has created the vision of Cognitive Radio. CRs are a new type of radio device, with the ability to detect and exploit residual bandwidth left idle by licensed users, under the condition that they do so without impacting on primary users. For a WMN, Cognitive Radio represents a way to improve overall utilization of available spectrum and expand the spectrum available to individual network[9].

The paper is presented in the following sequence. A introduction about the related work is presented in previous paragraphs. In section II, proposed work is presented. In section III, issues and technologies related to Cognitive Radio is presented. In section IV, routing protocols in wireless mesh networks are presented. In

section V, developed GUI's. Finally conclusion are presented in section VI.

II. PROPOSED WORK

With Cognitive Radio(CR) capabilities, we consider the capacity gains achieved by utilizing CR in the presence of multiple primary networks. In the Cognitive scenario, with CR enabled, rather than simply blocking flows ,the home network can search for another network to service the flow.The home network (K) borrows (lease) a channel by requesting another network (K') in order to service the flow.

III. COGNITIVE RADIO

A. Issues

A Cognitive Radio "is a radio that is aware of its surroundings and adapts intelligently" [6]. Early work in CR, led by Mitola, focused primarily on upper layer adaption in which the radio platform responded directly to anticipated user or application needs. The radio seeks out the required information and provides the user with instructions or the desired service[10].

Spectrum utilization can be improved significantly by allowing a secondary user to utilize a licensed band when the primary user is absent. Cognitive radio, an agile radio technology has been proposed for efficient use of spectrum. A CR is able to fill in, spectrum holes and serve its users without causing harmful interference to the licensed user by sensing and adapting. To do so, the CR must continuously sense the spectrum it is using in order to detect the reappearance of the primary user. Once the primary user is detected, the CR should withdraw from the spectrum so as to minimize the interference it may possibly cause. This is a very difficult task, as the various primary users will be employing different modulation schemes, data rates, and transmission powers in the presence of variable propagation environments and interferences generated by other secondary users.

B. Technologies

The bandwidth continues to increase; the users are seeking solutions for the apparent scarcity. The licensed spectrum is relatively unused across many time and frequency slots. To solve the problem of spectrum scarcity and spectrum under utilization, the use of CR technology is being considered.

In Haykin's paper[7], it was stated that a cognitive radio is an intelligent wireless communication system that is aware of its surrounding environment.

CR Technology will enable the users to [8]:

- Determine which portions of the spectrum are available and detect the presence of licensed users when a user operates in a licensed band.
- Select the best available channel.
- Coordinate occurs to this channel with other users.
- Vacate the channel when a licensed user is detected[11].

IV. ROUTING PROTOCOLS IN WMN

Only a few protocols have been developed specifically for WMNs. Several approaches have been considered. MIT(SrcRR) and MeshNetworks (MeshNetworks Scalable Routing) designed new protocols tailored for WMNs. MeshNetworks Scalable Routing (MSR) is a hybrid routing protocol, supposedly able to support highly mobile users and to dynamically adapt to networks conditions. As the protocol is not in the public domain, it is not possible to verify the company's claims. SrcRR is a variation of DSR using the expected transmission time as a metric instead of the number of hops. In other words, the shortest paths are determined based on least packet loss.

Other works have focused on enhancing existing routing protocols with new routing metrics more appropriate for WMNs. Indeed, the fixed wireless backbone allows a better estimation of the link quality through regular measurements. It is also possible to introduce channel diversity in the network infrastructure so as to reduce interference and increase overall throughput.

A. Spectrum Sensing

Spectrum sensing and environment awareness are the basis of cognitive behavior, it will map the system to a certain state according to the outside environment. Spectrum sensing gets the information of spectrum space.The spectrum sensing can be shown in Fig 1.

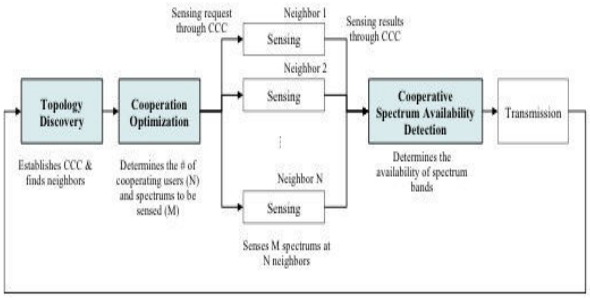


Fig. 1 : Spectrum Sensing

V. DEVEOPED GUIS

The simulation is done in the JAVA environment. Codes are written & when the code is run, after entering various inputs & other details, the simulation results are observed. The various graphical user interfaces obtained for showing the performance of Cognitive Radio based WMN is presented in the Figs. 2 to 15 respectively

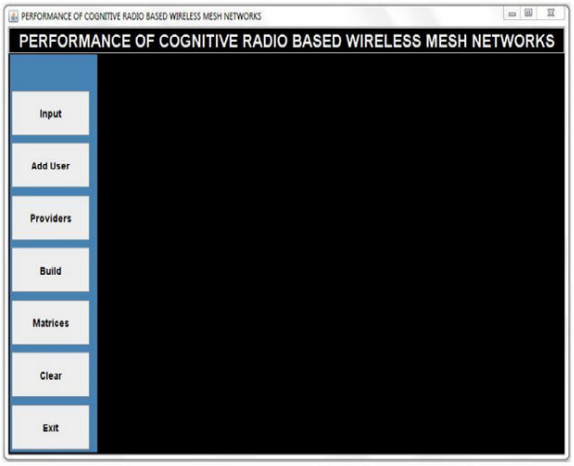


Fig. 2 : Home page of proposed system

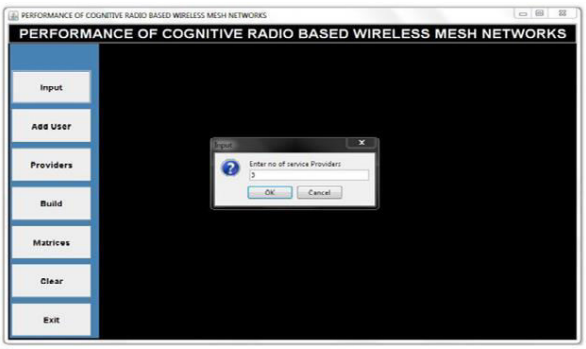


Fig. 3 : Entering no of service providers

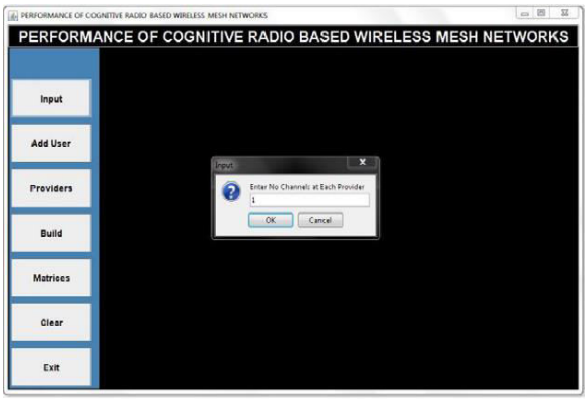


Fig. 4 : Enter no of channels at each provider

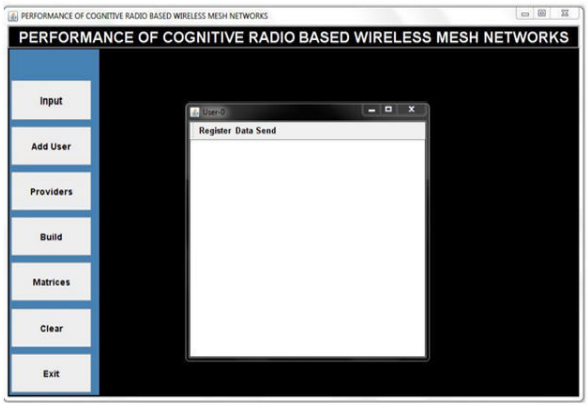


Fig. 5 : Creation of users

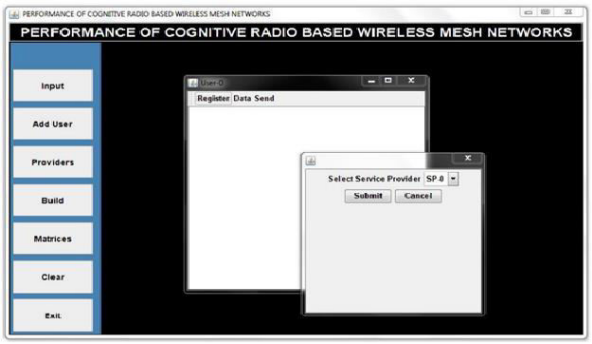


Fig. 6 Selecting Service Provider for given User

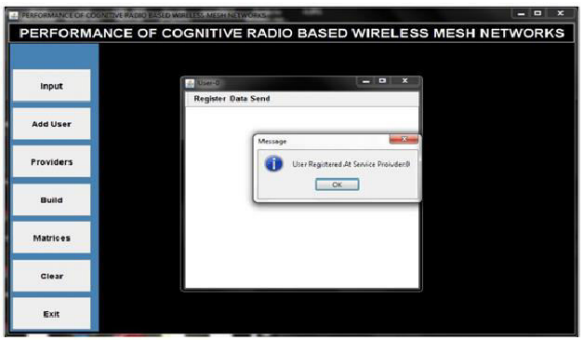


Fig. 7 : Registration of users to service providers

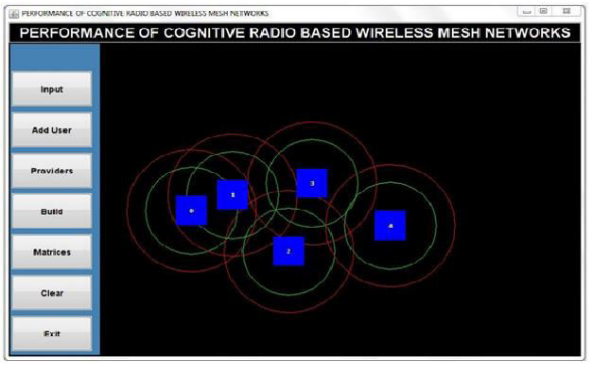


Fig. 8 : Creation of wireless mesh nodes

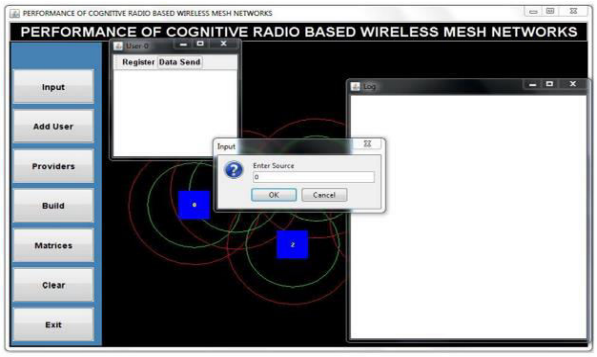


Fig. 9 : Enter source for sending data

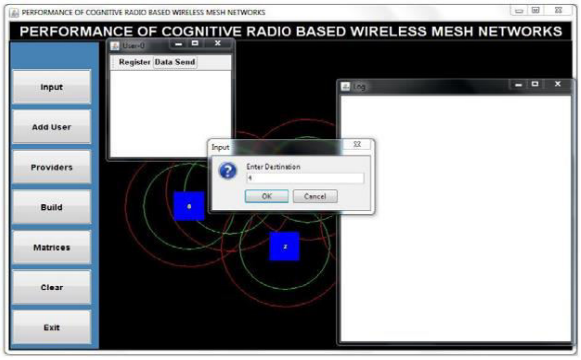


Fig. 10 : Entering destination for receiving data

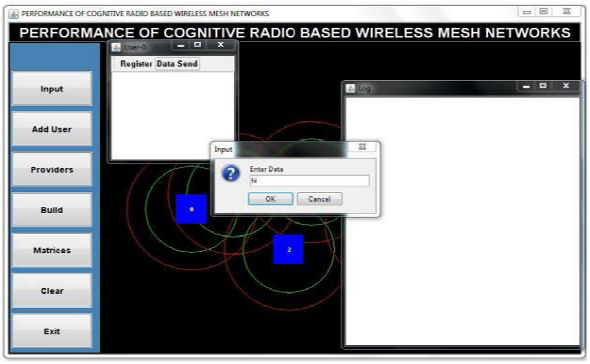


Fig. 11 : Enter data to send

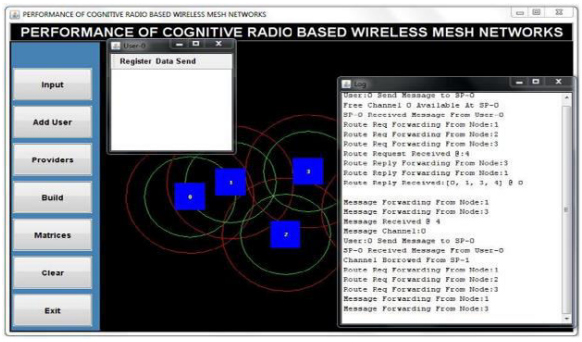


Fig. 12 : Displaying log file for route request & reply

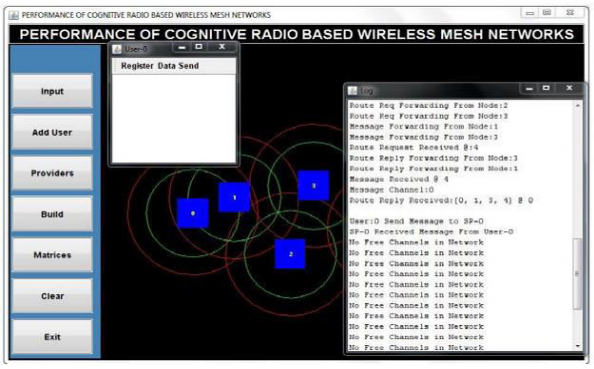


Fig. 13 : Log file showing no free channels available In network

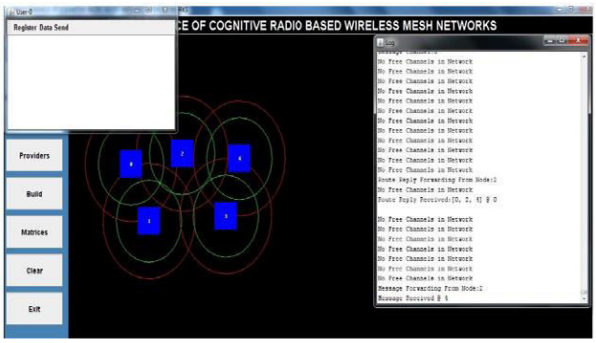


Fig. 14 : Log file showing available free channels

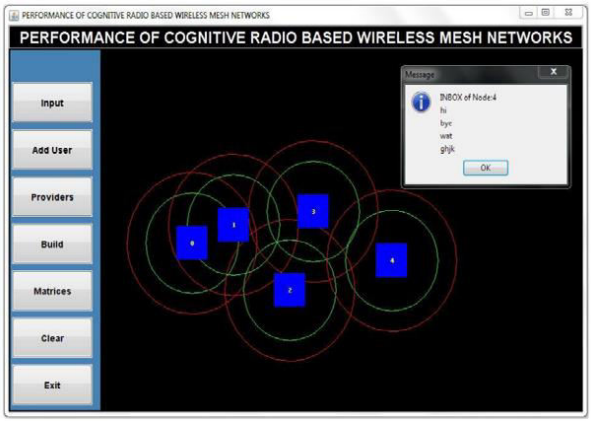


Fig. 15 : Illustration of messages displayed on receiver's inbox

VI. CONCLUSION

In this work, we have studied the improvements that can be gained from using CR in a WMN. From a classic system where primary service providers have assigned channel resources and all users belong to a home network (and must receive their service on their service provider's channels), we added CR-capabilities. CR-enabled nodes can operate on any channel, as

required. It was demonstrated that CR capabilities provide several benefits to a WMN.

First, CR improves the QoS that can be provided to network flows. Alternatively, more flows can be supported by a service provider, or fewer channel resources can be used. Second, CR increases utilization by allowing the network more flexibility to service flows, particularly longer (more hops) flows.

Third, CR provides a mechanism for QoS differentiation, where a CR-enabled achieves a higher level of service from the network, while still maintaining minimum service constraints for classic users.

REFERENCES

[1] MIT Roofnet, http://www.pdos.lcs.mit.edu/ roofnet, 2010.

[2] Seattle Wireless, http://www.seattlewireless.net, 2010.

[3] R. Draves, J. Padhye, and B. Zill, "Routing in Multi-Radio, MultiHop Wireless Mesh Networks," Proc. ACM MobiCom, Sept. 2004.

[4] R. Draves, J. Padhye, and B. Zill, "Comparison of Routing Metrics for Static Multi-Hop Wireless Networks," Proc. ACM SIGCOMM, Sept. 2004.

[5] FCC Spectrum Force Report, Policy Task Federal Communication Commission (USA), ET Docket No. 02-135, Nov. 2002.

[6] J. Mitola, "Cognitive radio: Model-based competence for software radios," Ph.D. dissertation, Dept. of Teleinformatics, KTH, 1999.

[7] S. Haykin, "Cognitive radio: Brain-empowered wireless communications," IEEE J. Sel. Areas Commun., vol. 23, pp. 201–220, Feb. 2005.

[8] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey," Comput. Netw., vol. 50, pp. 2127–2159, 2006.

[9] Nizar Bouabdallah, "Performance of Cognitive Radio-Based Wireless Mesh Networks," IEEE on Mobile Computing, vol. 10, no.1, January 2011.

[10] Allen B. MacKenzie, "Cognitive Radio and Networking Reasearch at Virginia Tech," IEEE proceedings, vol. 97, no. 4, April 2009.

[11] Khaled Ben Letaief, "Cooperative Communications for Cognitive Radio Networks," IEEE proceedings, vol 97, no. 5, April 2009.

◈◈◈

Graphical Visualization of Risk Assessment for Effective Risk Management during Software Development Process

Raghavi K Bhujang [1] & Suma. V[2]

[1]Dept of MCA, PESIT, Bangalore & Member, Research and Industry Incubation Centre, RIIC, DSI, Bangalore
[2]Research and Industry Incubation Center, Dayananda Sagar Institutions, Bangalore, India
E-mail : raghavi@pes.edu, sumavdsce@gmail.com

Abstract - Success of any IT industry depends on the success rate of their projects, which in turn depends on several factors such as cost, time, and availability of resources. These factors formulate the risk areas, which needs to be addressed in a proactive way. The rudimentary objective of risk management is to circumvent the possibility of their occurrence by identifying the risks, preparing the contingency plans and mitigation plans in order to reduce the consequences of the risks. Hence, effective risk management becomes one of the imperative challenges in any organization, which if deemed in an apt way assures the continued sustainability of the organization in the high-end competitive environment. This paper provides visualization of risk assessment through a graphical model. Further, the matrix representation of the risk assessment aids the project personnel to identify all the risks, comprehend their frequency and probability of their occurrence. In addition, the graphical model enables one to analyze the impact of identified risks and henceforth to assign their priorities. This mode of representation of risk assessment factors helps the organization in accurate prediction of success rate of the project.

Keywords - *Risk Management, Risk Assessment, Software Quality, Software Development Process*.

I. INTRODUCTION

The survivability of any industry is driven by the successful nature of projects that are engineered on board. However, the success of any project depends on several influencing areas, which require consistent and periodical attention by the project manager. Some of these areas are schedule management, finance management, change management, conflict management, etc. It is worth to note that each of the aforementioned areas emerges as risk if not managed in righteous way. Risk is a chance for occurrence of an unlikely event, which would result with highly unacceptable consequences. Thus, risk management is one of the critical challenges that need to be addressed in a skillful and efficient manner. It is therefore necessary to take preventive measures to reduce the likelihood of risk occurrences and to minimize its impact in realizing effective project management.

The process of risk management starts with identification of risks and classifying them into different types. Subsequently, every classified risk is assessed with impact level, its probability of occurrence and frequency of occurrence that enables one to prioritize according to their severity. The systematic analysis of risk aids project personnel in achieving an accurate predictive estimation of the apt choice of process and resources in the project.

State of the art in the risk management domain indicates progress of research in risk associated issues. Authors of [1] recommend the classification of every identified risk in a risk based tree structure in order to assess and resolve them efficiently. They suggest probabilistic calculation approach for qualitative and quantitative analysis and assessment of these risks. However, classification of risk should include not only the type and it is essential to consider the probability and frequency of occurrence of every identified risk. This facilitates to realize the impact of the risks, which results in formulation of prioritization list.

This paper aims to bring in a crispy classification of risk based on several modulating factors such as risk type, risk probability and risk frequency, which are deemed three independent factors. Consequently, establishing dependency relation between these factors assists to analyze the risk impact efficiently using matrix representation.

The organization of this paper is as follows, section II of this paper provides details of the related work in the domain of risk management. Section III elucidates the analysis of the investigation. Section IV provides

matrix representation and mathematical analysis of the risk assessment. The last section of this paper provides the summery of this investigation.

II. LITERATURE SURVEY

Since the evolution of software engineering, risk management has become one of the key challenges in day-to-day software development activities. Authors of [2] have introduced a systematic and qualitative project analysis of risk using Risk Factor Analysis (RFA) method. They state that RFA technique enables risk analysts to develop projects effectively.

Study made by authors of [3] on data collected over several industries indicates the extent of usage of risk management practices in industries. They suggest the impact of risk management to depend mainly on better meeting time and budget goals and less on product performance and specification. However, authors of [4] focus on the benefits of implementing effective risk management tools and techniques in software development project.

Further, authors in [5] analyzed the risks involved during software design. They feel that business goals determine risks and risk drives methods while methods yield measurement resulting in measurement to further drive decision support. Consequently, they state that decision support drives fix/rework and application quality.

On the other hand, authors of [6] have developed a tool named Risk Failure Mode and Effects Analysis (RFMEA) which is an extension of the Failure Mode and Effects Analysis technique. The benefit of the RFMEA include an increased focus on the most imminent risks, prioritizing risk contingency planning, improved team participation in the risk management process, and development of improved risk controls.

Authors of [7] have investigated the various risk sources in design and build projects quality using the discriminant analysis technique where risk is classified into one of the three risk groups namely cost, time and quality. However, authors of [8] provide an explanation on what enterprise risk management is and how an operational risk management fits into the ERM framework.

III. RISK ASSESSMENT

Risk is the consequence of inadequate information. Every identified risk is assessed for its classification. Classification of risks further enables one to evaluate the vitality of the risk in the effective development process. It is further elucidated below reflecting the significant activities to be formulated in order to develop high quality risk-free deliverable.

A. Risk Identification

Identifying the risks is not a component of a single area or a phase in software development and is a requisite activity involved in the entire project. This process involves all personnel who will be affected by the risks such as customers, stakeholders, Subject Matter Experts (SME), Risk management experts, project managers, team members and end users. Process areas, which fall into the uncertainty level causes threat and hence deemed as risk areas in project management. Thus, all the uncertainty areas need to be focussed to identify the risk factors, which are either internal or external risk factors and are estimated from the development level up to the organisational level. This process of risk identification begins at the project conceiving point, up to the project deployment point as risk can crop up at any point of time and in any phase within a project, which needs to be monitored periodically.

In lieu of avoiding the risks from popping up, risks need to be welcomed within a project by which a mitigation plan can be prepared which lead to the reduction of the level of uncertainty in future projects. With identification of every risk (irrespective of the phase of the project), it is categorized in to different risk types depending on the area that gets the impact and further based on the probability of occurrence of the risk in addition to consideration of frequency of its occurrence.

B. Risk Classification

With the assessment of every identified risk, it is a wise justification to further classify the risks. Our wide spectrum of investigation on several projects across various software industries has resulted in the classification of risk to be depending on risk type, on probability of risk occurrence, on frequency of occurrence of the risk, on risk impact and on risk priority.

TABLE I. depicts the sample risk type classification which is shown at a higher level of abstraction.

TABLE I : SAMPLE RISK TYPES AT THE COARSE LEVEL OF ABSTRACTION

Risk Type	**Risk Type Description**
Technology Risk	The Software / Technology related to the project may have risks.
Cost Risk	Risk associated with the ability of the project to achieve the planned life-cycle costs
Schedule Risk	Risk associated with the adequacy of the time allotted for the planning, R & D,

	facility design, construction, and startup operations.
Scope	Risk that comes when Requirements are ignored for the sake of technology.
People	Project lacks enough staff or those with the right skills.
Requirements	Project changes are managed poorly.
Estimation	The time required to develop the software is underestimated.
Tools	The code generated by CASE tools is inefficient. CASE tools cannot be integrated.
Organizational	The organization is restructured so that different management are responsible for the project.

TABLE I. infers the existence of several types of risks at a coarse level of abstraction.

Risk probability is the probability of occurrence of risk that normally varies from 0% occurrence to 100% occurrence where the boundary values of 0 and 100 will never be met. TABLE II. illustrates the probable encounter of risk in the development process. As an instance, the probability of occurrence of risk is 0% indicates the non-occurrence of the risk. However, the probability of risk occurrence being 100% is a case of certainty and hence is not considered under the probability classification of risk. TABLE II. infers that any type of risk, varying between 0 % and 100% is only considered for risk mitigation and further processing.

TABLE II : PROBABILITY OF RISK OCCURRENCE

Risk Type	**Probability of occurrence**
Breach of organizational and project standards	0%
Defects	100%
All risks types as mentioned in Table 1.	0<x<100 where x represents the percentage of probability of risk occurrence

Since, not all risks can be avoided even after the implementation of a mitigation plan, the frequency of occurrence of these risks can take up different values before implementing this plan and after implementing the same. For instance, in a microwave application the performance of a timer is considered as a risk factor. Frequency of the timer not functioning may be 7 times in an hour and hence mitigation plan is formed to resolve the risk. With the implementation of the mitigation plan, the risk occurrence frequency would have come down to 2 times in an hour. TABLE III. illustrates the frequency of risk occurrence for the sample risk types.

TABLE III : FREQUENCY OF RISK OCCURRENCE FOR SAMPLE RISK TYPES

Risk Type	**Risk Frequency (probable frequency)**	**Risk Description**
Requirements	Frequent	Likely to occur very often and/or continuously.
Schedule Risk	Likely	Occurs several times over the course of a transformation cycle.
People	Occasional	Occurs sporadically.
Technology Risk	Seldom	Remotely possible and would probably occur not more than once in the course of a transformation cycle.
Change in Organizational Standards	Unlikely	Will probably not occur during the course of a transformation cycle.

TABLE III. infers a sample of risk types and its estimated frequency of occurrence. The frequency of occurrence of risks in project management is estimated using historical data regarding the number of times it has occurred until then and its subsequent occurrences.

IV. GRAPHICAL VISUALIZATION OF RISK CLASSIFICATION

Risk management is one of the core needs of the day. From this investigation, it is apparent that analysis of risk is influenced by several factors such as risk type, probability of risk occurrence, frequency of occurrence of specified risk. However, our study has thrown light on the fact that risk is independently influenced by the aforementioned modulating factors while the other risk influencing factors such as impact and priority of the risk are dependent on these three above-stated factors. Since the existence of dependency relations on risk influencing factors, it can be visualized in a mathematical model using graphical representation and matrix representational scheme. The mathematical visualization of risk factors enables one to accurately estimate and predict the efficiency of every deliverable during the software development process, which does currently not exist in the industrial atmosphere. Figure

1. depicts the graphical representation of risk assessment factors.

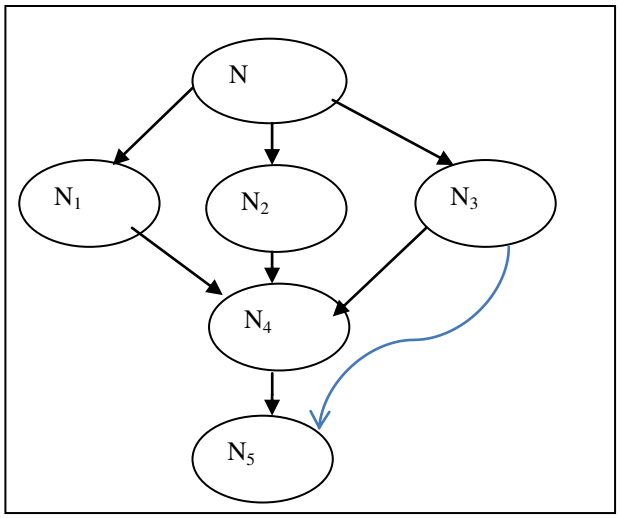


Fig. 1 : Graphical representation of risk assessment factors

Where, N – Risk

N_1 – Risk Type

N_2 – Risk Probability

N_3 –Risk Frequency

N_4 – Risk Impact

N_5 – Risk Priority

And (N_1,N_2, N_3) are the independent factors.

For any Risk (N), assessment of risk is made by considering its type, probability of occurrence and frequency (N_1, N_2, N_3). However, they further influence N_4, which in turn influences N_5. Hence, (N_1, N_2, N_3) influences N and influences N_4 while N_4 influences N_5. Thus, there exists a transitive relation, R, between the risk assessment factors. Figure 2. Figure 3. and Figure 4. represents aforementioned analysis .

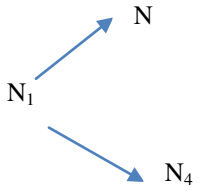


Fig. 2 : Assessment of Risk type

Figure 2 infers that R (N_1) = {N, N_4} i.e. Risk Type influences both Risk and the Risk Impact.

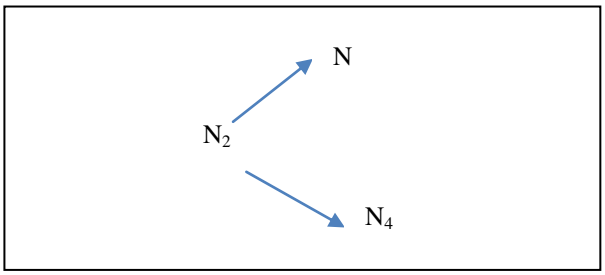


Fig. 3 : Assessment of Risk type

Figure 3 infers that R (N_2) = {N, N_4} i.e. Risk Probability also influences both Risk and the Risk Impact.

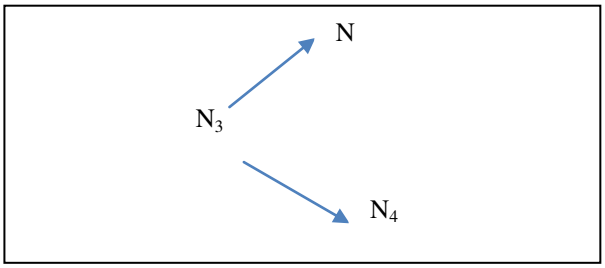


Fig. 4 : Assessment of Risk type

Figure 4 infers that R (N_3) = {N, N_4} i.e. Risk Frequency also influences both Risk and the Risk Impact. However,

N4 ⟶ N5 i.e. R(N_4) = {N_5} i.e. Risk Impact influences Risk Priority.

Where these R (N_1)...R (N_4) are the relational sets and N_1,N_2 ,N_3 are the independent factors. These independent factors can take up different values depending on the type of project.

Figure 5. depicts a matrix representation of the above analysis.

$$M_R = \begin{array}{c|cccccc} & N & N1 & N2 & N3 & N4 & N5 \\ N & 0 & 0 & 0 & 0 & 0 & 0 \\ N1 & 1 & 0 & 0 & 0 & 0 & 0 \\ N2 & 1 & 0 & 0 & 0 & 0 & 0 \\ N3 & 1 & 0 & 0 & 0 & 0 & 0 \\ N4 & 0 & 0 & 0 & 0 & 0 & 1 \\ N5 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}$$

Fig. 5 : Matrix representation of risk assessment factors

Figure 5. shows that $N_1 \rightarrow N_4 \rightarrow N_5$ i.e. N_1 (Risk Type) influences N_4 (Risk Impact) which in turn influences N_5(Risk Priority).

This is a transitive relation which says that $N_1 \rightarrow N_5$., $N_2 \rightarrow N_5$ and also $N_3 \rightarrow N_5$

It is worth to note that, this can be further proved by applying FloydWarshall's Algorithm on matrix M_R. Floyd Warshall's algorithm is used for finding the transitive closure of a relation R. It takes as input the matrix M_R representing the relation and output the matrix M_{R*}

$$M_R^* = \begin{array}{c} \\ N \\ N1 \\ N2 \\ N3 \\ N4 \\ N5 \end{array} \begin{array}{c} \begin{array}{cccccc} N & N1 & N2 & N3 & N4 & N5 \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Where the values of column N_5 for N_1, N_2,N_3 are replaced by the new transitive values i.e.1.

This algorithm will be beneficial when the order of the matrix is high i.e. when more number of factors are involved which needs to be identified with new relationships.

The focus of this paper is to provide a graphical visualization of risk assessment factors and to provide a mathematical analysis of the relations existing between them. However, our forthcoming papers explore the impact analysis of these representations.

V. CONCLUSION

Sustainability of any organization depends on the effective risk management capability of the company. Ever since the existence of industries, successful risk management has become one of the fundamental activities on board.

This investigation on several projects across various software industries has indicated that existing risk management operates on domain knowledge and experience of resource personnel. However, mathematical perspective of risk management is more result oriented than supposition oriented. This paper therefore focuses on graphical visualization of risk assessment, which enables the resource personnel to accurately estimate and predict the risk factors. Nevertheless, the mathematical representation is valid for known and identified risks.

REFERENCES

[1] Hooman Hoodat and Hassan Rashidi, "Classification and Analysis of Risks in Software Engineering", World Academy of Science, Engineering and Technology 56, 2009.

[2] John P Kindinger and John L Darby, "Risk Factor Analysis – A New Qualitative Risk Management Tool", Proceedings of the Project Management Institute Annual Seminars & Symposium, September 7–16, 2000 • Houston, Texas, USA.

[3] Tzvi Raz, Aaron J.Shenhar and Dov Dvir, "Risk Management, Project Success and Technological Uncertainty", R & D Management, 32, 2. Blackwell Publishers, 2002.

[4] Y.H. Kwak a, J. Stoddard, "Project risk management: lessons learned from software development environment", Technovation 24 (2004) 915–920.

[5] Denis Verdon, Gary McGraw, "Risk Analysis in Software Design", PUBLISHED BY THE IEEE COMPUTER SOCIETY. Also visit http://www.cigital.com/papers/download/bsi3-risk.pdf.

[6] Thomas A. Carbone, Donald D. Tippett, "Project Risk Management Using the Project Risk FMEA", Engineering Management Journal Vol. 16 No. 4 December 2004.

[7] O. E. Ogunsanmi1, O. A. Salako2, and O. M. Ajayi3, "Risk Classification Model for Design and Build Projects", Journal of Engineering, Project, and Production Management 2011, 1(1), 46-60.

[8] Jeffrey Lee Cowherd and Daniel P. Manson, "Risk Management". Also visit, http://www.csupomona.edu/~jis/2003/Cowherd_Manson.pdf

◈◈◈

Performance Comparison in MANETs

Tejal Arvind Sonawale & Shikha Nema

EXTC, V.E.S.I.T, Chembur, Mumbai, India
E-mail : tejal10sonawale@gmail.com, aishuup@gmail.com

Abstract **-** Ad Hoc Networks face a lot of problems due to issues like mobility, power level, load of the network, bandwidth constraints, dynamic topology which lead to link breaks, node break down and increase in overhead. As nodes are changing their position consistently, routes are rapidly being disturbed, thereby generating route errors and new route discoveries. The need for mobility awareness is widely proclaimed. In our dissertation we present a scheme AOMDV-APLP that makes AOMDV aware of accessibility of neighbor nodes in the network. Nodes acquire the accessibility information of other nodes through routine routing operations and keep in their routing table. Based on this information route discovery is restricted to only "accessible" and "start" nodes. Further route with the strongest signal strength is selected from multiple routes using Link life value predicted by Link Breakage prediction technique. Simulation result shows that using accessibility and link life knowledge in route discovery process MAC overhead, routing overhead and average delay is reduced **3** times, and improve the Packet delivery ratio to a large extent than standard AOMDV which reflects effective use of network resources.

Keywords **-** *Ad hoc networks; Routing protocols; ; QoS,link breakage, accessibility prediction.*

I. INTRODUCTION

Wireless technologies are unequivocally among the most rapidly progressing technology sectors. There is a vast range of wireless technologies, applications and devices, which are either already a substantial part of our daily life or could play this role in future. Wireless ad hoc networking is one of these applications, which can potentially enhance our abilities to solve real life challenges.

Wireless ad hoc networking or Infrastructure-less networking can be considered as an extension to the autonomy that was anticipated with the introduction of wireless networking. Wireless ad hoc networking makes those real life scenarios possible where there is a need of instantaneous and prompt communication. There is a widespread range of scenarios, from conventions or meetings with people quickly sharing information to the emergency search-and-rescue operations, where such networks are well suited. A wireless ad hoc network is a random collection of devices with radio transceivers that accompany each other without any prior infrastructure in a temporary manner to collaboratively accomplish a task.

- The participants i.e. the devices or the nodes can be stationary, mobile, or both, and they can join or leave the network as per their requirement. Similarly, wireless ad hoc networks have technically no geographical limitations on their size; a wireless ad hoc network can be as large as possible provided that all the nodes are able to communicate with each other, though the commonly available range is restricted from the body area to the local area. The concept of wireless ad hoc networking has numerous real life applications as it provides a simple, flexible, effortless, and instant approach to communicate in a cooperative scenario.

- Efficient communication
- Technological limitations
- Resource limitations
- Security
- Quality of service

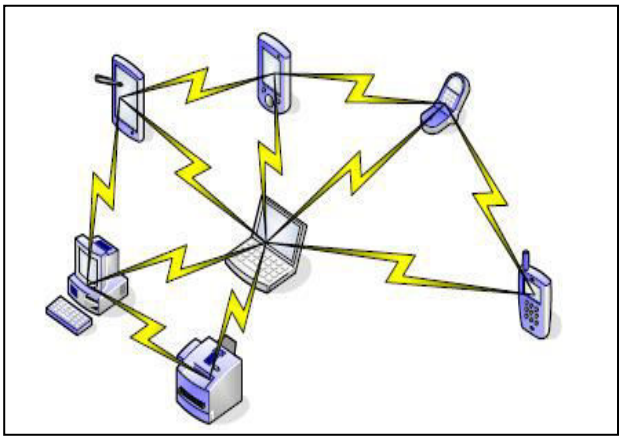


Fig.1: A typical wireless ad hoc networks

II. ROUTING PROTOCOL

AODV [11] is an improvement on DSDV. AODV makes use of the on-demand approach for finding routes. A route is established only when it is required by a source node for transmitting data packets and it maintains these routes as long as they are needed by the sources. AODV performs hop-by-hop routing by maintaining routing table entries at intermediate nodes. A node updates its route information only if the destination sequence number of the current received packet is greater than the destination sequence number stored at the node. This indicates freshness of the route and prevents multiple broadcast of the same packet. AODV makes use of the broadcast identifier number that ensures loop freedom since intermediate nodes only forward the first copy of the same packet and discard the duplicate copies. There are three phases of the AODV Routing Protocol. First is the Route Request, Route Reply and Route Maintenance phase. The Figure 2.1 displays a Wireless Ad Hoc scenario, which consists of 9 mobile nodes where the route has to be set from source (S) to destination (D).

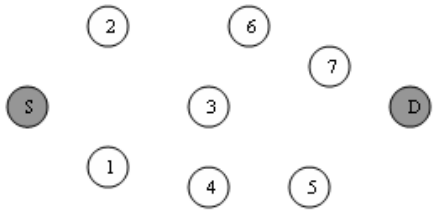


Fig. 2-1 : Wireless Ad Hoc Network Scenario

- **Route Request Phase:**

The route discovery process is initiated when a source needs route to a destination and it does not have a route in its routing table it floods the network with RREQ packet specifying the destination for which the route is requested.

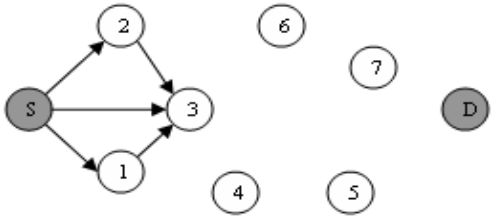


Fig. 2-2 : Route Request Broadcast

The figure 2.2 shows the broadcast of Route Request to the neighboring nodes. Here if node 3 has already received request from node S then it will discard the request that will come from node 1 and node 2. The nodes 1 and 2 will further broadcast it to their neighboring nodes 6 and 4 and if all the intermediate nodes do not have a route to the destination then the request is further broadcast to node 7 and 5 and thus it reaches the destination node D.

- **Route Reply Phase:**

The second phase is the Route Reply phase if the neighboring nodes have route to the destination then the node generates a RREP and sends back to the source along the reverse path and if it does not have the route then the request is forwarded to other nodes. Once the source node receives the RREP it can start using the route to send data packets. The source node rebroadcasts the RREQ if it does not receive a RREP before the timer expires, it attempts discovery up to maximum number of attempts or else aborts the session. It also makes a reverse route entry in its routing table and then forwards the packet. S starts sending the data from whichever route it receives the RREP and then changes the route if it receives the route with a less hop count.

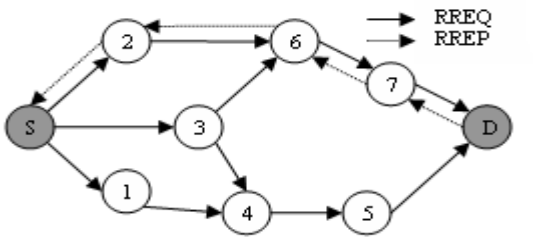


Fig. 2-3 : Route Reply Phase

The table 2.1 shows the routing table of AODV maintained by each node. The table consists of 5 fields the address of the destination node, sequence number, hop count, next hop and expiration time out. As each node just contains a single route to the destination if this route fails then a new route discovery has to be run by the source node. The Destination entry of the routing table specifies the node D where the source has destined the packet. The Sequence Number helps in maintains the freshness of the route. The Hop Count specifies the number of hops required by the source to reach the destination. The Next Hop specifies the next hop taken by the data to reach the destination D.

Table 2-1 Routing Table of AODV

Destination	Sequence Number	Hop Count	Next Hop	Expiration Timeout
D	1234	4	2	…..

- **Route Maintenance phase:**

If one of the intermediate nodes changes its position or fails then the neighboring node realizes the link failure and sends a link failure notification to its upstream neighbors. After the link failure notification

has reached the source it will reinitiate a route discovery if needed. The HELLO messages are sent at regular intervals by the intermediate nodes to find the correct information of the neighboring node. Here if the link between node 2 and node 6 goes down then a new route discovery is run and a path is set up between node S and node D.

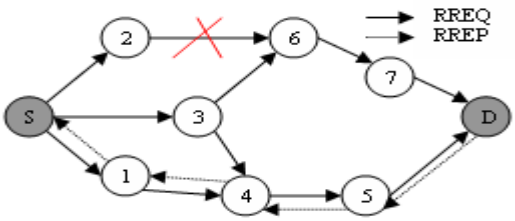


Fig. 2-4 : Route Maintenance Phase

The routing table2.2 displays the modified table of AODV protocol after the link failure.

Table 2-2 Modified Routing Table of AODV

Destination	Sequence Number	Hop Count	Next Hop	Expiration Timeout
D	1234	4	1	…..

The biggest drawback of AODV is with respect to its route maintenance. If a node detects a broken link while attempting to forward the packet to the next hop then it generates a RERR packet that is sent to all sources using the broken link. The source runs a new route discovery after receiving RERR packet. The frequent route breaks cause intermediate nodes to drop packets because no alternate path to destination is available. This reduces overall throughput, packet delivery ratio and increases average end-to-end delay if there is high mobility. The other drawback is that multiple RREP packets are received in response to a single RREQ packet and can lead to heavy control overhead. The HELLO message leads to unnecessary bandwidth consumption. Let us have a look at the already existing AODV protocol.

AOMDV

Adhoc On-demand Multi-path Distance Vector (AOMDV) [12] is an extension to the AODV. The main difference lies in the number of routes found in each route discovery. A little additional overhead is required for the computation of multiple paths. This protocol does not require any special type of control packets but makes use of AODV control packets with a few extra fields in the packet headers. The AOMDV protocol computes multiple loop-free and link-disjoint paths. There are three phases of the AOMDV protocol. The first phase is the Route Request, second is the Route Reply and the third phase is the Route Maintenance phase.

- **Route Request:**

The protocol propagates RREQ from source towards the destination. The figure 2.5 will show the working of AOMDV, which allows multiple RREQ to propagate. The node S as shown in Figure 2.5 has to set a path to the destination node D. So node S as in AODV broadcasts multiple requests to its neighboring nodes 1 and 2. This means that request with same sequence numbers are sent to the destination node. They further broadcast the request to the other neighboring nodes, which are further sent to the destination node D.

- **Route Reply:**

The protocol establishes multiple reverse paths both at intermediate nodes as well as destination. Multiple RREPs traverse these reverse paths back to form multiple forward paths to the destination at the source and intermediate nodes. If the intermediate nodes have the route defined for the destination then they send the RREP to the source node S. The protocol is designed to keep track of multiple routes where the routing entries for each destination contain a list of next hops together with the corresponding hop counts. All the hop counts have the same sequence number then the path with the minimum hop count is selected and all the other paths are discarded. The protocol computes multiple loop-free and link-disjoint paths. Loop-freedom is guaranteed by using a notion of “advertised hop count”. Each duplicate route advertisement received by a node defines an alternative path to the destination. To ensure loop freedom, a node only accepts an alternative path to the destination if it has a lower hop count than the advertised hop count for that destination. The advertised hop count is generally the maximum hop count value possible for a node S to reach a node D. If any value that is received by the source S is greater than the advertised hop count value then a loop is formed so this RREP is discarded. The multiple RREPs are received by the source via multiple paths and a minimum hop count route is selected, the other routes carrying a higher hop count value are discarded.

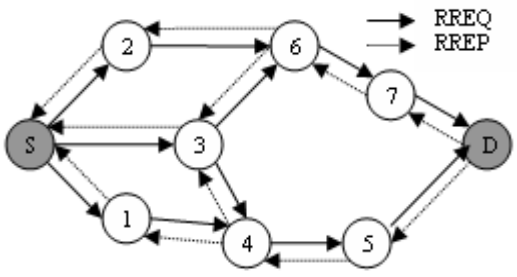


Fig. 2-5 : Working of AOMDV

Destination is the node where the packet is destined to, the sequence number to maintain the freshness of the routes, the advertised hop count that avoids the formation of loops. The route list consists of Hop Count required to reach a particular destination, Next Hop is the next hop the packet is supposed to take to reach the required destination, Last Hop is the last hop taken to reach the destination. If the packet is following the same path then this value is same as the Next Hop or else it changes and Expiration Timeout is the time for which the path will exist. There are multiple entries for a single destination but the routes that contain the lowest hop count are only recorded in the routing table and the other routes are discarded.

- **Route Maintenance Phase:**

The third phase is the Route Maintenance Phase. This phase works in exactly same as AODV. If the intermediate nodes are not able to receive a response of the HELLO message then they broadcast a Route Error message. After receiving this message all the nodes that use the particular route to reach the destination make this particular route as infinity and inform the source node to run a fresh route discovery. The routing table after a link break will appear as follows:

As node 3 has gone down the modified routing table of S will appear as above. When node 7 or node 5 goes down and there are no routes left in the routing table of S then the route discovery will be run. So it surely provides an improvement over AODV.

The above mechanism establishes loop free paths at every node but these paths have to be made disjoint. There are two types of disjoint paths, one is the node disjoint and the other is the link disjoint. Node-disjoint paths do not have any nodes in common, except the source and destination. The link disjoint paths do not have any common link.

An AODV protocol is been developed which develops route on-demand. The biggest drawback of AODV is with respect to its route maintenance. If a node detects a broken link while attempting to forward the packet to the next hop then it generates a RERR packet that is sent to all sources using the broken link. The source runs a new route discovery after receiving RERR packet. The frequent route breaks cause intermediate nodes to drop packets because no alternate path to destination is available. This reduces overall throughput, packet delivery ratio and increases average end-to-end delay if there is high mobility. The other drawback is that multiple RREP packets are received in response to a single RREQ packet and can lead to heavy control overhead. The HELLO message leads to unnecessary bandwidth consumption.

The AOMDV is an extension to the AODV protocol for computing multiple loop-free and link-disjoint paths. The protocol computes multiple loop-free and link-disjoint paths. Loop-freedom is guaranteed by using a notion of “advertised hop count”. Each duplicate route advertisement received by a node defines an alternative path to the destination. To ensure loop freedom, a node only accepts an alternative path to the destination if it has a lower hop count than the advertised hop count for that destination. With multiple redundant paths available, the protocol switches routes to a different path when an earlier path fails. Thus a new route discovery is avoided. Route discovery is initiated only when all paths to a specific destination fail. For efficiency, only link disjoint paths are computed so that the paths fail independently of each other.

In AOMDV RREQs reaching the node may not be from disjoint paths, if RREQ is from one common node one of the RREQ is discarded, this messages implicitly provide knowledge about the mobility and accessibility of their sender and originator. for example, if node A is constantly receiving messages initiated by another node B, this implies that node B is relatively stationary to node A. furthermore a valid route from node A to node B is available either directly or through other nodes. Instead of discarding repeated RREQs messages node can perform additional computation on available routing data and predict accessibility of other nodes. In terms of cost, AOMDV-AP has two additional characteristics. Firstly, repeated RREQs are used for routing table maintenance. Certainly, the additional overhead of performing this action is negligible because this RREQ is already available to the routing agent and all it has to do is to update one or two entries in the routing table. Secondly, routing entries remain permanently in the routing table. As a result, routing tables have more entries (and they also have an additional field in every entry). Use of repeated RREQs further stimulates this issue by adding entries, which were usually discarded. However, in our view, for an ad hoc network with a fair number of nodes such a situation will not cause serious problems. Larger routing tables have a positive role too. During the route discovery process, intermediate nodes can generate RREPs if they have a valid route to the destination; thereby, flooding of RREQ is obstructed. Undoubtedly, flooding has the worst effects on the performance of an ad hoc network.

Now AOMDV[8] routing make use of pre-computed routes determined during route discovery. These solutions, however, suffer during high mobility because the alternate paths are not actively maintained. Hence, precisely when needed, the routes are often broken. To overcome this problem, we will go for link breakage prediction. Prediction will be done only for multiple paths that are formed during the route

discovery process. All the paths are maintained by means of periodic update packets unicast along each path. These update packets are MAC frames which gives the transmitted and received power from which distance can be measured. This distance can be used to predict whether the node is moving inward or outward relative to the previous distance value that is it give the signal strength. At any point of time, only the path with the strongest signal strength is used for data transmission.

III. AOMDV WITH ACCESSIBILITY PREDICTION

In AOMDV repeated RREQs are not discarded. All duplicate RREQs arriving at the node are examined but not propagated further as each duplicate defines an alternate route. Thus AOMDV allows for multiple routes to same destination sequence no. With multiple redundant paths available, the protocol switches routes to a different path when an earlier path fails. Thus a new route discovery is avoided. Route discovery is initiated only when all paths to a specific destination fail. Routing table entry has one common expiration timeout regardless of no of paths to the destination. If none of the paths are used until the timeout expires, then all the paths are invalidated and the advertised hop count is reinitialized. While doing all this, routing information such as RREQs,RREP and REER packets collected can be used to predict the accessibility of nodes. This prediction is used to reduce routing overhead, MAC overhead and to enhance packet delivery ratio and connection success ratio.

3.1. Accessibility Prediction algorithm

a) If a node A receives a routing packet from another node B, node B is in A's neighborhood and is accessible to A.

b) If a node A receives a routing packet originated by a node B, node B is accessible to node A and there exists a valid route from node A to node B.

c) If a node A receives a RERR from a node B, all the unreachable nodes mentioned in this RERR are no more accessible to node A through node D.

Routing entries will never be deleted a new field "Accessible" is added to each routing table entry depicts the predicted accessibility information possible values

Start = No information

Accessible = A valid route to node exists or would be possible

Inaccessible = A valid route to node would not be possible

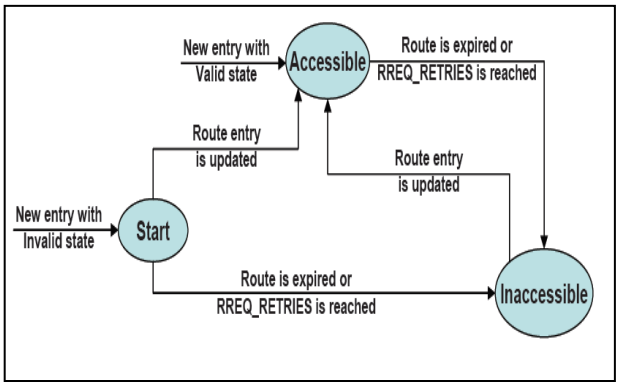


Fig: 3.1 State diagram of AOMDV with accessibility Prediction

Table 3-1 : Routing Table of AOMDV-APLP

Destination
Sequence number
Advertised_hopcount
Expiration _timeout
Route list {(nexthop1,hopcount1),(nexthop2,hopcount2),…
Accessibility

COST

No extra messaging is required.

Additional computation due to "Accessibility" field is negligible.

Computation cost of using repeated RREQs is negligible.

Routing table entries are never deleted.

Size of routing table might not be a problem in a reasonable size network.

Relative stationary nodes are good candidate to be included in route

3.2 Modified route discovery with accessibility prediction

There is no route discovery for "Inaccessible" nodes, which reduces overhead. The value of the accessibility field is just a prediction. It is likely that this information gets stale. To assume an "Inaccessible" node "Accessible" is not an issue as in such a situation usual AOMDV procedures will be followed. However, the converse could have serious consequences. For example, nodes can conserve plenty of resources by not performing route discoveries for "Inaccessible" nodes, provided the prediction is correct. However, if this prediction is incorrect, this resource conservation will cost them in the form of connectivity loss and

consequently throughput loss. Thus, in such a situation there is a trade-off between overhead reduction (or resource conservation) and connectivity (or throughput).

IV. MODIFIED AOMDV WITH LINK BREAKAGE PREDICTION

Now AOMDV with accessibility prediction routing protocol make use of pre-computed routes determined during route discovery. These solutions, however, suffer during high mobility because the alternate paths are not actively maintained. Hence, precisely when needed, the routes are often broken. To overcome this problem, we will go for link breakage prediction. Prediction will be done only for multiple paths that are formed during the route discovery process. All the paths are maintained by means of periodic update packets unicast along each path. These update packets are MAC frames which gives the transmitted and received power from which distance can be measured, this distance can be used to predict whether the node is moving inward or outward relative to the previous distance value that is it give the signal strength. At any point of time, only the path with the strongest signal strength is used for data transmission. Following is the method to calculate link lifetime.

From two ray ground model we get Transmitted power PT and Received power PR using which we can calculate distance 'd' by given formula.

P_r = **k** $\frac{P_t}{d^4}$ **where k =** $G_t.G_r.(h_t.h_r)^2$ **is a constant**

A link breakage algorithm is used to predict the value of $\mathbf{t_{break}}$ using **'d'.**

4.1 Link Breakage Algorithm

Now $\mathbf{t_{break}}$ can be calculated by the following algorithm

Always assume nodes moving radially outward. Initially

$$V = V_{prev} = V_{\max}\ m/s, d_{prev} = 0.0m$$

$$v = \left| \frac{d - d_{prev}}{t - t_{prev}} \right|$$

$$V = (w)^* v + (1 - w)^* V_{prev}\ \text{k}$$

w based on ratio of time since last sample (!t = t – t prev) and average sample interval T

Time dependency of w ensures quick adaptation to change

$$t_{break} = \left[\frac{d_{\max} - d}{V} \right]$$

$$V_{prev} = V; d_{prev} = d$$

Thus Accessibility and Link Breakage Prediction (APLP) techniques are implemented in AOMDV protocol the proposed protocol has produced good results. The proposed AOMDV-APLP protocol has reduced MAC overhead, Routing overhead and end-to-end delay. On account of which packet delivery ratio is increased a lot as compared to AODV-AP (accessibility prediction) and standard AOMDV.

V. PERFORMANCE METRICS

- **MAC overhead** – the total number of all kinds of MAC packets generated during the simulation time. The retransmission of data frames are also included in it.
- **Routing overhead** – it includes all kinds of AOMDV packets generated as well as forwarded during simulation.
- **Average Delay** – The average end–to-end delay is defined packets traveling from the source to the destination node. The packets generally sometimes get delayed due to transmission, processing, collision and queuing.
- **Packet Delivery Ratio** – The ratio of total number of data packets successfully received by all the destinations to the total number of data packets generated by all the sources.

VI. CONCLUSION

AODV came up with the advantage of the routes being discovered a single route on-demand but this caused a lot of packet delay, Routing and MAC overhead on node failure as a new route discovery had to be run by the source and RREQs are send to all the nodes. AODV-APLP came up with the solution of above problem but the number of routes to the destination is one. AOMDV came up with the advantage of multiple routes being discovered and the route carrying the minimum hop count value is selected but it suffers from large Routing, MAC overhead and Packet delay on node failure, because RREQs are send to all the nodes neighboring nodes. We proposed and implemented AOMDV-APLP where RREQs or route discovery is initiated only for "Accessible" and 'start" nodes which reduces the MAC overhead, Routing overhead, Packet Delay.

Results show that, our proposed protocol, reduces packet delay by 70%, and increases packet delivery ratio

considerably as compared to standard AOMDV protocol. Our protocol also gives stable connectivity as route with the strongest signal strength is selected with the help of Link lifetime.

REFERENCES

[1] www.ietf.org/html.charters/manet-charter.html

[2] Murthy, S. and J.J. Garcia-Luna-Aceves, An Efficient Routing Protocol for Wireless Networks, ACM Mobile Networks and App. J., Special Issue on Routing in Mobile Communication Networks, Oct. 1996, pp. 183-97.

[3] C. E. Perkins and P. Bhagwat, "Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers", in Proceedings of ACM SIGCOMM 1994, August 1994, pp. 234-244.

[4] G. Pei, M. Gerla and T.-W. Chen, Fisheye State Routing in Mobile Ad Hoc Networks. In Proceedings of the 2000 ICDCS Workshops, Taipei, Taiwan, Apr. 2000, pp. D71-D78

[5] B. Johnson, D.A. Maltz, Y.-C. Hu, "The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks", Internet-Draft, draft-ietf-manet-dsr-10.txt, July 2004,

[6] V. Park, and S. Corson, Temporally Ordered Routing Algorithm (TORA) Version 1 Functional Specification. IETF Internet draft, 1997.

[7] Y. B. Ko and N. H. Vaidya. Location Aid Routing (LAR) in mobile ad hoc networks. In Proc. ACM/IEEE MOBICOM, Oct. 1998.

[8] Z. J. Haas and M.R Pearlman, "The Zone Routing Protocol (ZRP) for ad hoc networks", IETF Internet draft , August 1998.

[9] Navid Nikaein, Christian Bonnet and Neda Nikaein, "HARP - Hybrid Ad Hoc Routing Protocol", in proceeding of IST 2001: International Symposium on Telecommunications, Iran/Tehran 2001.

[10] M. Joa-Ng and I-Tai Lu, "A peer-to-peer zone-based two-level link state routing for mobile ad hoc net-works", IEEE on Selected Areas in Communications, vol. 17, no. 8, pp. 1415 1425, 1999.

[11] C. E. Perkins and E. M. Royer, "Ad Hoc On-Demand Distance Vector Routing", Proceedings of IEEE Workshop on Mobile Computing Systems and Applications 1999, February 1999, pp. 90-100.

[12] Mahesh K. Marina and Samir R. Das,"On-Demand Multipath Distance Vector Routing in Ad Hoc Networks", in IEEE Proceedings 2001.

[13] C. E. Perkins, E. M. Royer, S. R. Das, and M. K. Marina,"Performance comparison of two on-demand routing protocols for ad hoc networks",. IEEE Personal Commun. Mag., vol. 8, pp. 16.28, Feb. 2001.

[14] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network

[15] routing protocols", in Proceeding. Int. Conf. on Mobile Computing and Networking, Oct. 1998.

[16] C. E. Perkins, ed., Ad Hoc Networking, ch. 3. Addison Wesley, 2001.

[17] Rohit Dube, Cynthia D. Rais, Kuang-Yeh Wang, and Satish K. Tripathi, "Signal stability based adaptive routing (SSA) for adhoc mobile networks", in IEEE Personal Communication, February 1997, vol. 4.

[18] Ian D. Chakeres and Elizabeth M. Belding-Royer, "The utility of hello messages for determining link connectivity", in Proceedings of the 5th International Symposium on Wireless Personal Multimedia Communications (WPMC), Honolulu, Hawaii, October 2002.

[19] T. Camp, J. Boleng, V. Davies, "A survey of mobility for ad hoc network research", in Proc. WCMC, 2002, pp.483-502.

[20] Michael Bahr, "Proposed Routing for IEEE 802.11s WLAN Mesh Networks", In Proc. 2nd Annual International Wireless Internet Conference, 2006.

[21] P Sambasivam, A Murthy, EM Belding," Dynamically Adaptive Multipath Routing based on AODV", In Proc. 3rd Annual, 2004- Citeseer

[22] Rehman, H. Wolf, L. "Performance enhancement in AODV with accessibility prediction", in mobile adhoc and sensor system, MASS 2007 IEEE International conference,pp 1-6

[23] p://www.isi.edu/nsnam/ns-build.html#allinone

◈◈◈

Customizalbe Keyboard using a Simple Paper

Ganesh Aithal P N, Megha M Kamath, Manjunath B & Shashank S

Atria Institute of Technology, Bangalore– 560024
*(Affiliated to **Visvesvaraya** Technological University, Belgaum)*

Abstract - With fast growing technology, all devices are becoming compact & there is a human feeling that there should be some simple system which helps to enter text, type long mails etc on our PDA, Pocket PC, and Mobiles etc. The standardized keyboard is difficult for people with disabilities to use, and the bacteria which lives in the keyboard are not suitable for medical environments. To solve this problem, we give a new insight into the characteristics of CMOS camera and microcomputer, and use them to improve a kind of customizable keyboard based on some prior accomplishments. In this paper, dominate features and expected applications of CMOS camera will be presented. This kind of customizable paper keyboard will be widespread in medical environment, transportation, and future computers.

Keywords - *CMOS Camera; AVR Microcomputer; Virtual Keyboard.*

I. INTRODUCTION

Today world is in race of miniaturization; Cell phones, PDA, Pocket PC etc day by day all getting smaller and smaller, but our hand and fingers cannot do the same. And for this problem we have come forward with a perfect solution. In this paper we are describing an emerging technology, which replaces the bulky keyboards with virtual keyboards. These keyboards are based on CMOS image camera. The CMOS camera continuously captures images of the region where the printed keyboard is placed and checks these images for finger placements and special image processing algorithms are used to recognize the finger and its position which is directly related to key**.** The camera watches finger movements and translates them into keystrokes in the device which can be your Palm Pc, Mobile Phone, PDA etc.

This paper keyboard can be placed on any flat surface, such as desktops, airplane tray tables, kitchen counters, etc. and can theoretically be interfaced with any computing device that requires text entry. This would eliminate the need to carry anything around and also prevent any chance of mechanical damage to the keypad in harsh environments.

II. BLOCK DIAGRAM

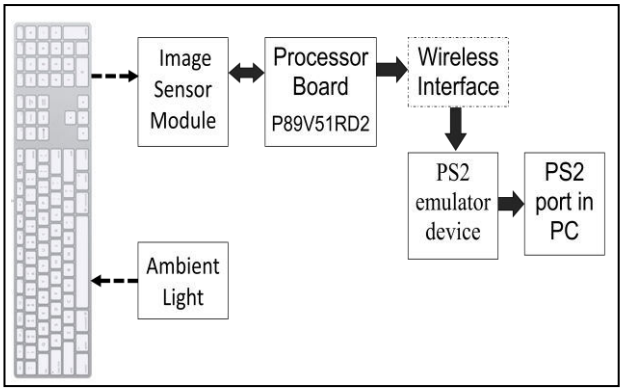


Fig 1: Generalized Block Diagram

III. FEATURE OF IMAGE SENSOR

Along with the CCD image sensor, CMOS have excellent imaging performance. CMOS image sensor has advantages such as low power consumption, low fabrication cost, compatibility with VLSI integration, the possibility of smaller system size and radiation hardness. The disadvantages of CMOS (Complementary metal oxide-semiconductor) image sensors, such as low-sensitivity, high noise, dark current and low filling degree have been resolved by technologies like DRSCAN noise[2] cancellation technology and it is now

widely used in surveillance, aviation and detection equipment, medical equipment, eye mask recognition and visual communications

IV. IMPLEMENTATION

Based on the characteristics of the CMOS camera, we improved and implemented a new kind of virtual keyboard by referring to some relative kinds of virtual keyboards. In this part we will discuss the hardware implementation from the following five aspects

A. Overall Descriptions

The Virtual Keyboard mainly contains four units, which are AVR Microcomputer, CMOS Camera, PS2 emulator device and the printed keyboard paper. The four units connect to each other and make up the whole project as follows.

B. CMOS Camera

In the project, we use Omni Vision's CMOS image sensor OV7640/8 as our capturing images device, because it is relatively cheaper and it can output image color data in progressive scan mode. Progressive scanning function is a crucial factor as we do not have enough power available to process the entire frames at once.

C. AVR Microcomputer

ATmega32 is a high-performance, low-power AVR 8-bit Microcontroller. It works above 5V power resources with 0-16MHz speed levels. We uses COM PORTA on Atmega32to communicate with computers and the computer offers power to the microcomputer through USB. Of course, AVR Microcomputers with higher performance is welcomed; besides, we can also choose double microcomputers to enhance its processing speed.

D. Customized Keyboard

The design of the keyboard is the uniqueness of the device, because user can design it according to our minds. No matter you prefer rectangular or circular, it's up to you, and the size or the content of the button are also highly personalized. It will change ways of inputting texts dramatically. We designed three types of customized keyboard as paragons, and also, the traditional printed one which is better for us to do accuracy test. It is highly suggested that the material of the keyboard surface should not be reflective.

E. Casing the Hardware

The camera of the device is designed at a fixed position such that it overlooks the printed keyboard from a particular angle and it can cast all the appropriate region of the keyboard. The camera above detects the signal and sends keystroke's position to the microcomputer to determine the input content.

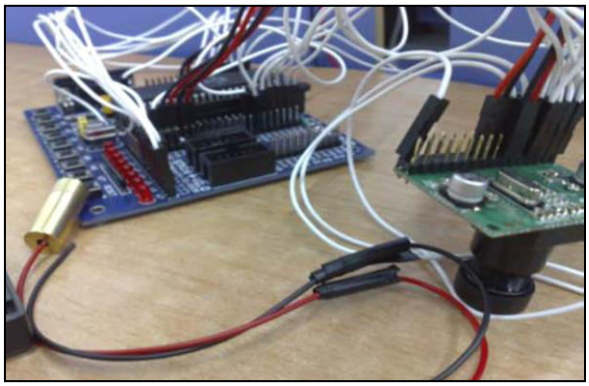


Fig 2 : OV7640/8 Camera Connecting to ATmega32 Microcomputer

V. WORKING

The layout of our bulky keyboard is initially printed on a simple paper. The general set up of our paper keyboard is, the printed paper is kept on the desk, over which, at a small distance from the paper, the µcam is been placed. Rests of the modules are then connected between the PS2 port of the PC and µcam.

When the user just poses his finger as if he is pressing the key, the camera captures the image. An important aspect of the camera is that it continuously captures the image. The captured image is not been saved in any external memory. Instead, it is been processed in real time. Hence the required storage capacity and the time required for the processor to process the image is less.

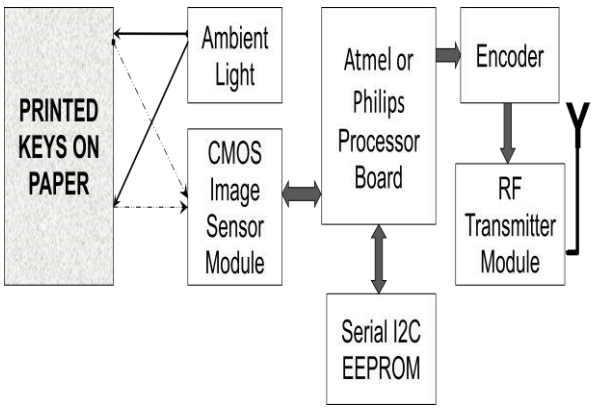


Fig 3(a): PS2 Transmitter Region

There will be a look-up table which contains the make & break codes of each key of the keyboard and the position of the key on the keyboard. Depending on that position, the make code is extracted from the table and sent to the PS2 emulator through RF transmitter receiver pair.

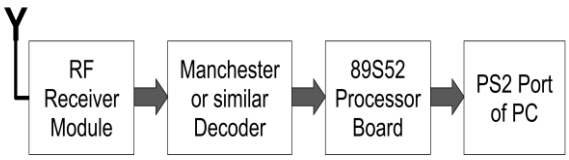


Fig 3 (b): PS2 Receiver Region

The received data is further decoded to obtain its originality and from the obtained code the PS2 emulator, acting as the input device of PC instead of keyboard, sends the key to the PC. Recognizing the key, the PC displays it on the monitor.

VI. TESTING

The operation of such a process is tested or monitored using as LCD. The LCD is interfaced with the camera module to monitor the each process inside the camera. Starting from the synchronization till the size of the image is been displayed on the LCD. The Camera has few built-in codes for synchronization and other processes. Five main instructions used in our approach are as follows:

1. Synchronization: AA0D 00 00 00 00h
2. Initial: AA01 00 01 01 03h
3. Get Picture: AA04 02 00 00 00h
4. Data: AA0A 02 XX XX XXh
5. Acknowledgement: AA0E XX 00 00 00h

The first four MSB of the codes are fixed and standardized by the manufacturer.

VII. APPLICATIONS

Paper keyboard can be the main input device for Computers, PDA, Pocket PC, and Mobiles
This keyboard is noise free it does not make even slightest of noise and in scientific labs; where total noise free environment is required ,there our keyboard will be the ultimate option.

Is can be used with Smart phones, PDAs, email, word processing and spreadsheet tasks, Gaming control, TV remote control,

As our keyboard is touch free, it is very useful in dangerous places where a smallest spark due to key press can cause explosion.

VIII. CONCLUSION

Paper Keyboard uses sensor technology and artificial intelligence to let users work on any surface as if it were a keyboard. These key boards claim to provide the convenience of compactness with the advantages of a full-blown QWERTY keyboard. The Paper Keyboard is designed for anyone who's become frustrated with trying to put information into a handheld but doesn't want to carry a notebook computer around.

Thus Paper keyboards will make typing easier, faster, and almost a pleasure. This paper is served as the instigator to illustrate the application of CMOS camera and AVR Microcomputer in virtual keyboard. We hope this paper could make some contributions to the fast developing computer industry.

IX. ACKNOWLEDGEMENT

We would like to acknowledge and extend our heartfelt gratitude to the following persons who have made the completion of this thesis possible: Ms.Shilpa M N (Lecturer) of Atria Institute of Technology, Bangalore.

Particular appreciation goes to Mr. Nikhil S Arora not only for the countless hours of rough draft reading as it was in process, inspiring suggestions, encouragement, and great patience with my seemingly endless revisions and questions, but also for the immeasurable intellectual input and lively discussion throughout the duration of this project.

I am also grateful cordially to my friend Sujeeth, who has contributed enormously to my understanding of microcomputer, camera and image processing. Without a doubt, they are the most gifted and natural friends I have ever known.

REFERENCES

[1] Zhongxiang Zhao, Zhihai Xu, Meijuan Bian, Analysis of Technological Features and Types of Digital Camera [J/OL].CAMERA MAGAZINE. 2001, (7).

[2] Xudong Wang, Yutang Ye, Comparative research and future tendency between CMOS and CCD image sensor [J/OL].Electronic Design Engineering. 2010, 18(11).

[3] Naweed Paya, Venkat Ganesh. Customizable Virtual Keyboard[EB/OL] . [2010-04-12].
http://courses.cit.cornell.edu/ee476/FinalProjects/s2008/nap33_vsg3/nap33_vsg3/index.html

[4] Li Pan, Qin Zheng. “Thumb Controlled Key Extraction Method of Virtual Keyboard for Touch-Screen Mobile”, Computer Systems &Applications, November2010.

[5] Cheng-Huei Yang, Li-Yeh Chuang, Yu-Huei Cheng, Jun-YangChang, Cheng-Hong Yang. An Intelligent Chinese Phonetic On-Screen Virtual Keyboard System. [J/OL].2006, 18

[6] "Celluon Laserkey CL850 Bluetooth Projection Keyboard". [Retrieved 2010-03-31].
http://www.mobilefun.co.uk/celluon-laserkey-cl850-bluetooth-projection-keyboard-p15385.htm?referer=aff_aw_74960.

Efficient Handling of Orphan Nodes in Wireless Sensor Networks

Herald Antony. H & Anandkumar. K.M

Dept. of Computer Science and Engineering, Easwari Engineering College

Abstract - The unconnected nodes in the Wireless sensor networks are called as orphan nodes. These nodes occur due to failure in obtaining a network address from a router/parent device and also when a node has lost its connectivity with its parent also become an orphan node. The performance of most wireless sensor network applications degrades rapidly in the increase of orphan nodes. The orphan nodes cannot be negotiated because every node has its own importance. In this paper, the orphan problem can be divided into two categories: the orphan node may be a *full function device (FFD)* problem *i.e.* router capable node and *reduced function device (RFD)* problem *i.e.* end-device. If the router capable node is orphaned the children of that node are become orphaned. So it is preferable to detect and restore the orphan nodes present in the network. The aim of this paper is to improve the performance of the network by reducing the orphan nodes by restoring the orphan nodes. For this we use an address borrowing technique to use the unused address spaces remaining in the network. To avoid collision the number of devices and the depth of the network is restricted. The simulation results shows the proposed scheme reduce the orphan nodes.

I. INTRODUCTION

Wireless sensor networks are becoming popular due to simultaneous, high-speed sensing and data acquisition from multiple wireless sensor devices for both small scale and large scale applications which provides scalability, low power, synchronization, flexibility, robustness, security, size and cost. The emerging field of wireless sensor networks combines sensing, computation, and communication into a single tiny device. While the capabilities of any single device are minimal, the composition of hundreds of devices offers radical new technological possibilities. The power of wireless sensor networks lies in the ability to deploy large numbers of tiny nodes that assemble and configure themselves. Usage scenarios for these devices range from real-time monitoring of environmental conditions, ubiquitous computing environments, monitor the health care environments.

A Zigbee network consists of a base station (Coordinator), routers and end-devices and it supports Star, tree and mesh network topologies. In a star network, all devices have to directly connect to the coordinator. For tree and mesh networks, devices can communicate with each other in a multihop fashion.

A zigbee coordinator is responsible for initializing, maintaining and controlling the network. The base station initializes the network establishment by advertising child-request messages. The same procedure is repeated by all router-capable nodes after they obtain their parents. The reason for orphan node creation is mainly due to either the poor quality of the communication link during child-request advertising process or parent nodes already holding maximum number of children. The maximum number of children to a parent leads to low signal strength between parent and children. So the coordinator decides three parameters before forming a network: the maximum number of children of a router (*Cm*), the maximum number of child routers of a router (*Rm*), and the depth of the network (*Lm*). The coordinator and each router can have at most *Rm* child routers and at least *Cm – Rm* child end devices.

II. RELATED WORK

[3] A mathematical approach is used to restore the orphan nodes back to the network. Restoring orphaned nodes to the sensor network may require some compromises to be made since the parent nodes are restricted in the maximum number of children they can possess.

Optimal restoration can be achieved by finding the optimal parent node for each orphaned node. When a router-capable node becomes orphaned, it's all descendants become orphaned too. This orphan propagation was controlled by using a scheme called enhanced self configuration, in which the orphan node can spread its orphan state to all its descendants immediately, allowing its descendants to take appropriate action. An adaptive clustering technique was introduced to control the cluster size while limiting the number of orphan nodes. The salient feature of all these reported methods is to minimize the number of orphan nodes in the network

The main disadvantage of this method is an orphan node may keep attempting to reconnect with its previous parent node immediately after the disconnection from the network, thus wasting energy.

[2] In the orphan environment, the chance an orphaned device successfully recovers from all orphanings is very low (less than 4 %). The orphan problem in the Zigbee addresses assignment method. They insist that if one uses the random formation policy specified in Zigbee, the utilization of the address pool will be very low, and those devices that cannot receive network addresses will be isolated from the network and become orphan nodes. Re- association problem of an orphan node is presented to assess the adaptation potential of the set of available new parents, they utilize Parent Adaptation Quality Indicator (PAI). The main disadvantage of this method is even though the two kinds of solution they propose enable re-association with better parent, they do not address orphan propagation problem and also the address pool is very low

[1]The Orphan problems have been divided into two sub problems: the bounded-degree-and-depth tree formation (BDDTF) problem and the end-device maximum matching (EDMM) problem. BDDTF deals about connecting as many routers as possible to form a network and EDMM deals about connecting as many end devices as possible to the network.

They use three algorithms to solve the orphan problems. Centralized span & prune algorithm, Distributed depth then breadth search algorithm were used to solve the BDDTF problem and a Greedy and Probe phase was used to solve the EDMM problem.
From the above algorithms, the coordinator and routers can accept more routers and devices if they still have capacities. However, when a node cannot join the network because all its neighbors have run out of their address capacities, we say the node has become an orphan. But there are unused remaining address spaces in the network.

III. SYSTEM ARCHITECTURE

We use Distributed Borrowing Addressing (DIBA) scheme and orphan propagation scheme to solve orphan problem. DIBA is a method of borrowing addresses from the neighbour nodes and assign the borrowed address to the newly entering nodes in the network. The orphan propagation scheme is used notify the descendents of a router as they become orphan when the link between a router and its parent becomes failure

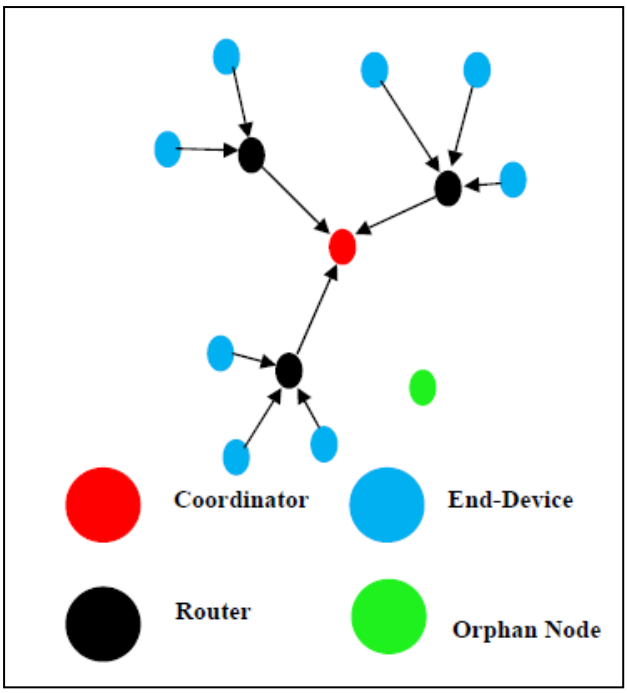


Fig 1. Overerall System Architecture

I. Distributed Addressing Scheme

All router nodes broadcast beacon frames at particular intervals. These frames contain the number of children they can add, that is the number of remaining address to assign. It is called as Available Address Count (AAC). If a new node enters into a network it searches for a parent by scanning. If the new node is located within coverage of more than one parent, it selects the node as its parent, which has more unoccupied addresses. Fig. 2 shows an example. The node *i* is located in the coverage of both nodes *m, c*. But the node *i* selects *c* as its parent instead of m because *m* is already holding 3 nodes *a, b, c.*

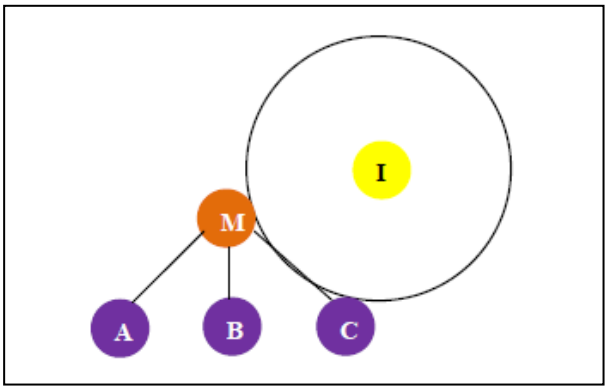
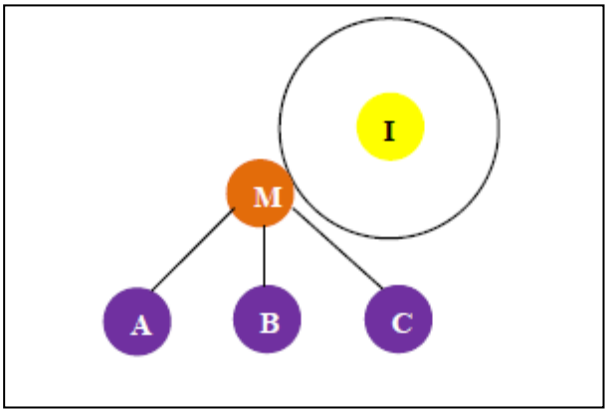


Fig. 2 : New Node entering a network

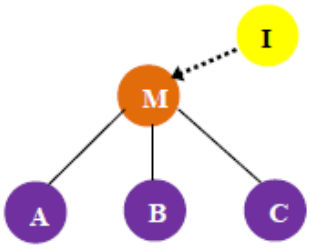
After scanning the new node will send Association REQ to the desired parent node. When the parent receives the request it will check its Available Address (AA). If the AA contains address it will allocate address to the new node. If the AA is zero and depth of the network *Lm* is more, the parent node obtains an address through the address-borrowing mechanism of DIBA. Figure 3 shows the operation of DIBA mechanism. For example the network has the following parameters, *Cm=3, Rm=3, Lm=3.*

The network has 4 nodes *m, a, b, c*. The node *m* is a parent for *a, b, c.* A new node *i,* enters into the network which is in the coverage range of node *m.* so the node *i* send an Association REQ to the node *m.* but the node *m* has a limit of 3 children (*Cm*=3), so node *m* cannot assign address to node *i.* To solve this, node *m* should broadcast an address borrowing request (AB_REQ) messages to borrow addresses from its descendent nodes, called neighbor nodes. The neighbor nodes *a*, *b*, and *c* respond to node *m* with address borrowing response (AB_RSP) messages including AA and AAC. If AAC is zero, no AB_RSP message is transmitted.
Node *m* sets a timeout value and checks the contents of the AB_RSP messages it receives from neighbor nodes during this interval. The node and address to borrow are selected with the following priority. First, an AA of the node with the biggest AAC is selected. If the AACs are dentical, an AA of the neighbor node with the highest address should be selected. If node *m* decides to borrow an address from node *c* through this process, as shown in Fig. 3, node *m* informs node *c* through the address borrowing ACK (AB_ACK) message. On receiving this message, node *c* records the address it has lent and updates its routing table. Node *m* makes node *i* its child by assigning the borrowed address through an association response message. Finally the newly entering node *i* joins the network.

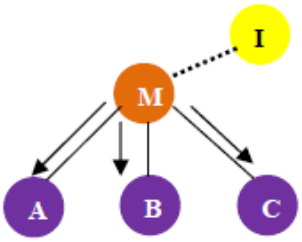
1. Scanning



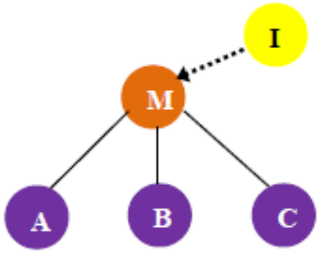
2. Association REQ



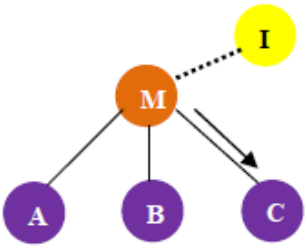
3. Address Borrowing REQ



4. Address Borrowing RSP



5. Address Borrowing ACK



6. Association RSP (Borrowed Address)

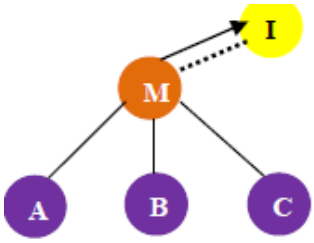


Fig. 3 : DIBA Scheme

III. ROUTING ALGORITHM

Fig. 3 shows an example WSN using DIBA. This network maintains a tree topology, and the parameters are Cm=3, Rm=3, Lm=4. The new node 6 wants to associate with node 0, but the AAC=0 because the node 0 already had 3 children. So the node 0 sends address borrowing REQ to its children. The descendent nodes reply the AA to the parent node 0. The parent node selects node 3 because it is having more AA than the other nodes. So the parent node replies an address borrowing ACK to the node 3. The node 0 borrows the address of node 3, and assigns the address to the node 6. Node 6 accepts nodes 7 and 8 as its children and assigns their addresses. Finally the node 6 is attached to the network.

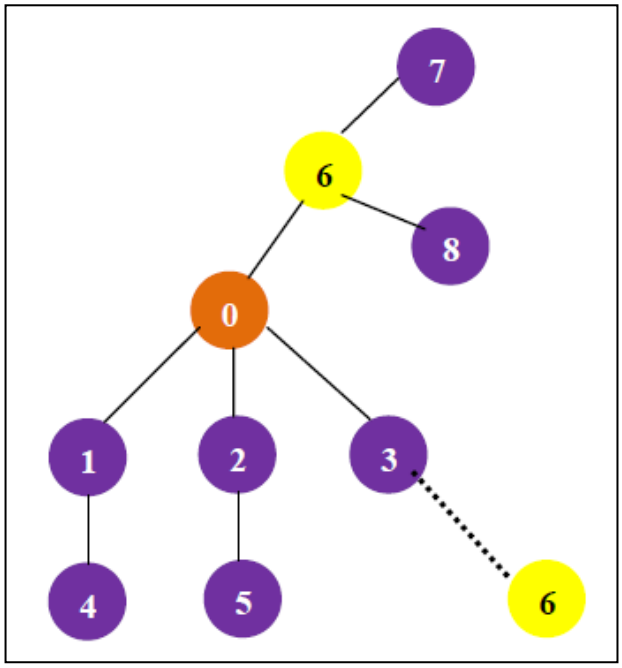


Fig. 4 : Example WSN using DIBA.

IV. SIMULATION & RESULTS

Fig 5(a) clearly shows the advantages of DIBA and generalizes the previous graphs. The top two lines show the average address assignment rates using DAAM and using DIBA according to the number of nodes in the network. This figure shows the address assignment rates when *C*m was in the range from 3 to 7. DIBA is superior to DAAM in the average address assignment rate. DIBA achieved address assignment rates 7% to 13% higher than those of DAAM. Because DAAM and DIBA guarantee address assignment rates from 30% to 70% in all cases, a 7% to 13% increase in address assignment rates is highly significant. The bottom line shows that IR averaged 21.92%, ranging from 18% to 24%. With random node placement (see Fig. 5(b)), DIBA achieved address assignment rates 7% to 8% higher than those of assignment rates were 40% to 55% in all cases, a 7% to 8% increase is meaningful. The IR averaged 21.92% when the case of 400 nodes was omitted, ranging from 6% to 23%.

Table 1. Routing tables of nodes shown in Fig. 4.

(a) Node 0's routing table

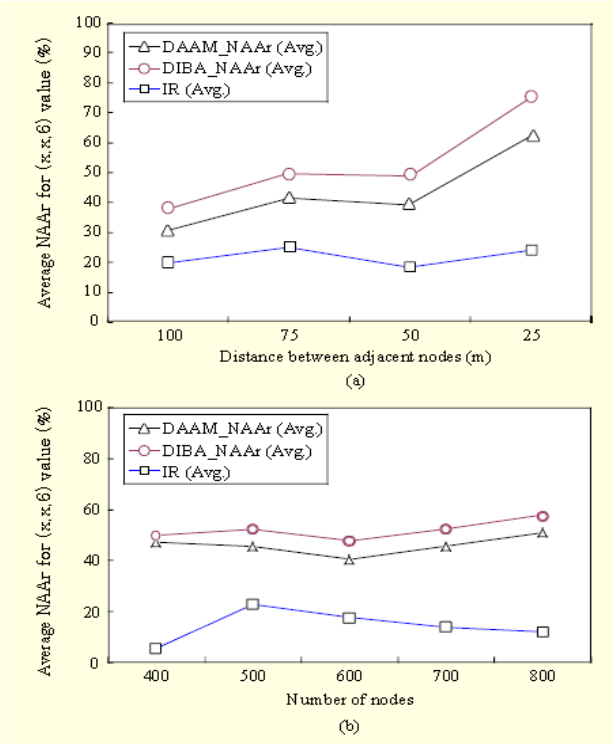
	Address	Lender	Borrower
Child 1	1	-	-
Child 2	2	-	-
Child 3	3	-	-
Child 4(Borrowed)	6	3	-

(b) Node 3's routing table

	Address	Lender	Borrower
Parent	0	-	-
Child 1	6	-	0

V. CONCLUSION

DIBA uses unused addresses from neighbour nodes. It has a routing algorithm to adapt the newly entering nodes. It solves the address assignment problem due to limited number of children. It uses the unused addresses in the network.



REFERENCES

[1] The Orphan Problem In Zigbee Wireless Networks, IEEE transactions 2009, Meng-Shiuan, Pan Chia-Hung, Tsai Yu-Chee Tseng

[2] Enhanced Self-Configuration Scheme for a Robust ZigBee-based Home Automation, IEEE transactions 2010, Kwang-il Hwang, Member, IEEE, Byoung-Jo Choi, Member, IEEE, Seok-hoon Kang

[3] Management of Orphaned-Nodes in Wireless Sensor Networks for Smart Irrigation Systems, IEEE transactions 2011, Suhinthan Maheswararajah, Saman K. Halgamuge, Member, IEEE, Kithsiri B. Dassanayake, and David Chapman

[4] IEEE TG 15.4b, Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs), IEEE Standard for Information Technology, 2006.

[5] ZigBee Alliance, ZigBee Specification Document 05347r17, Jan. 2008.

◈◈◈

Analysis of Customer Purchasing Transactions using Text Classification and Association rule mining

Vedula Venkateswara Rao & Koduri Sravan Kumar

Dept. of CSE, Sri Vasavi Engineering College, Pedatadepalli, Tadepalligudem—534101, India
E-mail: venkatvedula@yahoo.com, kforkoduri@gmail.com

Abstract **-** Decision Support is an important activity for retail marketing in large business organizations. Market Basket Analysis provides scientific decision support for retail market by mining association rules among different items involved in transactional database. In this paper we present a new approach to perform market basket analysis based on text classification and association rule mining. The text classification provides segmentation of item characteristics description to generate item's internal characteristics. Association rule mining then generates frequent item sets and association rules based on internal characteristics. This provides efficient decision support system. The method has been applied to Restaurant data and experimental results are validated.

Keywords **-** *Decision Support System, Market Basket Analysis, Association rule mining, frequent item sets, Hash Based Apriori, FPGrowth.*

I. INTRODUCTION

Mining Association rule is very important field of research in data mining. The problem of mining Association rule is put forward by R.S Agarwal first in 1993. Now the Association rules are widely applied in E-commerce, bank credit, shopping cart analysis, market analysis, fraud detection, and customer retention, to production control and science exploration etc. [1]. Mining results may not be interesting to people since they don't show semantic associations among items. However, the consequence may be different if association rules are mined on items' internal characteristics [1][2]. Consider a supermarket with a large collection of items. Typical business decisions that the management of the supermarket has to make include what to put on sale, how to design coupons, how to place merchandise on shelves in order to maximize the profit, etc. Analysis of past transaction data is a commonly used approach in order to improve the quality of such decisions. Until recently, however, only global data about the cumulative sales during some time period (a day, a week, a month, etc.) was available on the computer. Progress in bar-code technology has made it possible to store the so called basket data that stores items purchased on a per-transaction basis. Basket data type transactions do not necessarily consist of items bought together at the same point of time. It may consist of items bought by a customer over a period of time. Examples include monthly purchases by members of a book club or a music club.

This paper is organized as follows: Section II describes Basic Concepts of Association Rule Mining. Section III describes Proposed work. Section IV describes implementation details. Section V presents experimental results and performance analysis. Section VI presents conclusion and future scope.

II. BASIC CONCEPTS OF ASSOCIATION RULE MINING

Association rule finds interesting associations and/or correlation relationships among large set of data items. Association rule shows attribute value conditions that occur frequently together in a given dataset. A typical and widely-used example of association rule mining is Market Basket Analysis. For example, data are collected using bar-code scanners in supermarket. Such 'market basket' databases consist of a large number of transaction records. Each record lists all items bought by a customer on a single purchase transaction. Managers would be interested to know if certain groups of items are consistently purchased together. They could use this data for adjusting store layouts (placing items optimally with respect to each other), for cross-selling, for promotions, for catalog design and to identify customer segments based on buying patterns. Association rules do not represent any sort of causality or correlation between the two item sets The problem of mining association rules can be described as below: if $I = \{I_1, I_2, I_3, \text{-------}, I_n\}$ is the set of

items. Suppose D is database transaction set and each transaction T contains set of items, such that T⊆ I. Each transaction has identifier called as TID i.e. transaction id. Suppose A is a set of items and transaction T is said to contain A only if A ⊆T. Association rule is an implication like as A ⇒B i n which A, B ⊂ I and A ∩ B=Ø [6]. Definition of support: The support is the percentage of transactions that demonstrate the rule. An item set is called frequent if its support is equal or greater than an agreed upon minimal value the support threshold. [8].Definition of Confidence: Every association rule has a support and a confidence.

An association rule is of the form: X => Y.

X => Y: if someone buys X, he also buys Y.

The confidence is the conditional probability that, given X present in a transition, Y will also be present. Confidence measure, by definition:

Confidence(X=>Y) = support(X, Y)/ support(X)

The aim of association rule is to find all association problems having support and confidence not less than given threshold value. For the given support i.e. minsupp, if the item set of D‘s support is not less than minsupp, then it can say that D is the frequent item set.

The most common approach to find association rules is to break up the problem into 2 parts

1. Find Large Itemsets
2. Generate rule from the frequent Itemsets

A Large (Frequent) Itemset is an Itemset whose number of occurance is above the threshold (s).

III. PROPOSED WORK (ASSOCIATION RULE MINING AND MARKET BASKET ANALYSIS)

A. Goals of Market Basket Mining

Market Basket Analysis is a modelling technique based upon the theory that if you buy a certain group of items, you are more (or less) likely to buy another group of items. For example, if you are in an English pub and you buy a pint of beer and don't buy a bar meal, you are more likely to buy crisps (US. chips) at the same time than somebody who didn't buy beer. The set of items a customer buys is referred to as an item set, and market basket analysis seeks to find relationships between purchases. Typically the relationship will be in the form of a rule:

IF {beer, no bar meal} THEN {crisps}.

The probability that a customer will buy beer without a bar meal (i.e. that the antecedent is true) is referred to as the **support** for the rule. The conditional probability that a customer will purchase crisps is referred to as the **confidence**.

1. Association rules are statements of the form X1, X2, X3,……, Xn → Y meaning that if we find all of X1, X2, X3,……, Xn in the market basket then we have good chance of finding Y. The probability of finding Y for us to accept this rule is called the confidence of the rule. We normally would search only for rules that had confidence above a certain threshold. We may also ask that the confidence be significantly higher than it would be if items were placed at random into baskets. For example, we might find a rule like milk, butter → bread simply because a lot of people buy bread. However, the beer/diapers story asserts that the rule diapers → beer holds with confidence significantly greater than the fraction of baskets that contain beer.

2. Causality. Ideally, we would like to know that in an association rule the presence of X ,….,X causes" Y to be bought. However, —\causality" is an elusive concept. Nevertheless, for market-basket data, the following test suggests what causality means. If we lower the price of diapers and raise the price of beer, we can lure diaper buyers, who are more likely to pick up beer while in the store, thus covering our losses on the diapers. That strategy works because \diapers causes beer." However, working it the other way round, running a sale on beer and raising the price of diapers, will not result in beer buyers buying diapers in any great numbers, and we lose money.

3. Frequent Itemsets. In many (but not all) situations, we only care about association rules or causalities involving sets of items that appear frequently in baskets. For example, we cannot run a good marketing strategy involving items that no one buys any way. Thus, much data mining starts with the assumption that we only care about sets of items with high support; i.e., they appear together in many baskets. We then find association rules or causalities only involving a high-support set of items (i.e., X,…….,X → Y must appear in at least a certain percent of the baskets, called the support threshold.

Although Market Basket Analysis conjures up pictures of shopping carts and supermarket shoppers, it is important to realize that there are many other areas in which it can be applied. These include:

- Analysis of credit card purchases.
- Analysis of telephone calling patterns.
- Identification of fraudulent medical insurance claims.
- Analysis of telecom service purchases.

B. Frame Work of Frequent Item Set Mining

We use the term frequent Itemset for a set S that appears in at least fraction s of the baskets," where s is some chosen constant, typically 0.01 or 1%. We assume data is too large to fit in main memory. Either it is stored in a RDB, say as a relation Baskets (BID; item) or as a at _le of records of the form (BID; item1; item2;:::;itemn). When evaluating the running time of algorithms we:

- Count the number of passes through the data. Since the principal cost is often the time it takes to read data from disk, the number of times we need to read each datum is often the best measure of running time of the algorithm.

There is a key principle, called monotonicity or the a-priori trick that helps us find frequent Itemsets:

- If a set of items S is frequent (i.e., appears in at least fraction s of the baskets), then every subset of S is also frequent.

To find frequent Itemsets, we can:

1. Proceed level wise, finding first the frequent items (sets of size 1), then the frequent pairs, the frequent triples, etc. In our discussion, we concentrate on finding frequent pairs because:

(a) Often, pairs are enough.

(b) In many data sets, the hardest part is finding the pairs; proceeding to higher levels takes less time Than finding frequent pairs.

Level wise algorithms use one pass per level.

2. Find all maximal frequent Itemsets (i.e., sets S such that no proper superset of S is frequent) in one pass or a few passes.

C. Apriori Using Hashing

Our hash based Apriori implementation, uses a data structure that directly represents a hash table. This algorithm proposes overcoming some of the weaknesses of the Apriori algorithm by reducing the number of candidate k-Itemsets. In particular the 2-itemsets, since that is the key to improving performance. This algorithm uses a hash based technique to reduce the number of candidate Itemsets generated in the first pass. It is claimed that the number of Itemsets in C generated using hashing can be smalled,so that the scan required to determine L2 is more efficient.

For example, when scanning each transaction in the database to generate the frequent 1-itemsets,L1, from the candidate 1-itemsets in C1, we can generate all of the 2itemsets for each transaction, hash(i.e) map them into the different buckets of a hash table structure, and increase the corresponding bucket counts . A 2-itemset whose corresponding bucket count in the hash table is below the support threshold cannot be frequent and thus should be removed from the candidate set. Such a hash based apriori may substantially reduce the number of the candidate k-2 k+1 Itemsets examined.

Algorithm:

1. Scan all the transaction. Create possible 2-itemsets.
2. Let the Hash table of size 8.
3. For each bucket assign an candidate pairs using the ASCII values of the item sets.
4. Each bucket in the hash table has a count, which is increased by 1 each item an item set is hashed to that bucket.
5. If the bucket count is equal or above the minimum support count, the bit vector is set to 1. Otherwise it is set to 0.
6. The candidate pairs that hash to locations where the bit vector bit is not set are removed.
7. Modify the transaction database to include only these candidate pairs.

In this algorithm, each transaction counting all the 1itemsets. At the same time all the possible 2-itemsets in the current transaction are hashed to a hash table. It uses a hash table to reduce the number if candidate Itemsets. When the support count is established the algorithm determines the frequent Itemsets. It generates the candidate Itemsets as like the Apriori algorithm.

D. FP growth Algorithm

FPGrowth Algorithm allows frequent item set discovery without candidate item set generation. This algorithm generates frequent item sets from FP-tree by traversing in bottom-up fashion. The General idea is divide-and-conquer i.e Recursively grow frequent patterns using the FP-tree: looking for shorter ones recursively and then concatenating the suffix: –For each frequent item, construct its conditional pattern base, and then its conditional FP-tree; –Repeat the process on each newly created conditional FP-tree until the resulting FP-tree is empty. This the Algorithm is Two step approach:

Step 1: Build a compact data structure called the FP-tree Built using 2 passes over the data-set.

Step 2: Extracts frequent item sets directly from the FP-tree Traversal through FP-Tree.

Algorithm (FP-growth*: Mining frequent patterns with FP-tree by pattern fragment growth*)**.**

Input: A database *DB*, represented by FP-tree constructed according to Algorithm 1, and a minimum support threshold ξ.

Output: The complete set of frequent patterns.

Method: *call FP-growth*(FP-tree*, null*).

Procedure *FP-growth*(*Tree, α*)

{(1) *if Tree* contains a single prefix path // Mining single prefix-path FP-tree

(2) *then {* (3) *let P* be the single prefix-path part of *Tree*;

(4) *let Q* be the multipath part with the top branching node replaced by a *null* root;

(5) *for each* combination (denoted as *β*) of the nodes in the path *P do*

(6) *generate* pattern *β* ∪ *α* with *support* = *minimum support of nodes in β*;

(7) *let freq pattern set*(*P*) be the set of patterns so generated; *}* (8) *else let Q* be *Tree*;

(9) *for each* item *ai* in *Q do {* // Mining multipath FP-tree

(10) *generate* pattern *β* = *ai* ∪ *α* with *support* = *ai .support*;

(11) *construct β*'s conditional pattern-base and then *β*'s conditional FP-tree *Treeβ* ;

(12) *if Treeβ* = ∅ (13) *then call FP-growth*(*Treeβ, β*);

(14) *let freq pattern set*(*Q*) be the set of patterns so generated; *}* (15) *return*(*freq pattern set*(*P*) ∪ *freq pattern set*(*Q*) ∪ (*freq pattern set*(*P*) ×*freq pattern set*(*Q*)))

}

E. Max for Maximal Itemsets

A big problem of mining frequent item sets is that in many databases with long patterns, it would be computationally infeasible to enumerate all possible 2^k subsets of a frequent k-item set (k can easily be 30 or 40 or longer). Algorithms for mining frequent closed item sets are proposed since they are enough to generate association rules. However, **FCI** could also grow exponentially as **FI**. The set **MFI** is orders of magnitude smaller than the set **FCI**, and in many applications **MFI** is adequate to generate interesting patterns. FPMAX, which is an extension of the FP-growth algorithm, also finds the exact **MFI**. As with FP-growth, the highly compact FP-tree structure is used to store the information concerning frequent items. By adopting a pattern fragment growth method, it avoids costly candidate generation-and-test. A novel Maximal Frequent Item set tree (MFI-tree) structure is utilized to keep track of all maximal frequent item sets. This structure makes FPMAX perform subset checking more efficiently.

Based on the FP-growth algorithm, one can find all frequent item sets. But in order to solve our problem, some modifications are required to guarantee that the frequent item set generated by our algorithm is the longest frequent item set. We constructed a simple version by extending the FP-growth algorithm. When we constructed our variant of the FPMAX algorithm, we found that they are completely identical. It is not surprising, because the FPMAX algorithm is also an extension of the FP-growth algorithm. For consistency, we use a uniform name: FPMAX_LO ("Longest Only"). Like FP-growth, algorithm FPMAX_LO is recursive. The initial FP-tree constructed from the two scans of the database is passed on as the parameter of the first call of the algorithm. An item list *Head*, initialized to be empty, contains the items whose conditional FP-tree will be constructed from its conditional pattern base and will then be mined recursively. Before recursive call to FPMAX_LO, we already know that the combination set of *Head* and the items in the FP-tree is longer than the longest frequent item set found so far (guaranteed by line (7)). Thus if there is only one single path in the FP-tree, the items in this path, together with *Head*, constitute a longer frequent item set. If the FP-tree is not a single-path tree, then for each item in the header table, append the item to *Head*, construct the conditional pattern base of the new *Head*, and check in line (7) whether the combination set of *Head* with all frequent items *Tail* in the conditional pattern base is longer than the longest frequent item set so far. If yes, we construct the conditional FP-tree based on the conditional pattern base and explore this tree recursively.

Input: *T*: an FP-tree

Global:

lfi: the longest frequent item set found so far

Head: a list of items

Tree: the initial FP-tree

Output: The *lfi* that is a longest frequent item set

Method: Call **FPMAX_LO** (*Tree*).

Procedure **FPMAX_LO** (*T*) {

(1) IF *T* only contains a single path *P*

(2) THEN update *lfi* with *Head* ∪ *P*;

(3) ELSE FOR EACH item *i* in header table of *T* DO

{

(4) Append *i* to *Head*;

(5) Construct *Head*'s conditional pattern base;

(6) *Tail* = {frequent items in *Head*'s conditional pattern base};

(7) IF Length(*Head* ∪ *Tail*) > Length(*lfi*)

(8) THEN {

(9) Construct *Head*'s conditional FP-tree *THead*;

(10) **FPMAX_LO** (*THead*); }

(11) Remove *i* from *Head*. } //end of for each

} // end of procedure

IV. IMPLEMENTATION DETAILS

In this section, we analyze the performance of our ZBDD-based technique for mining frequent item sets by comparing with incremental FP Tree algorithm and traditional FP Tree algorithm. The algorithms were implemented in java language. Swing framework is used for designing GUI. We have placed transactional data records in data sets. The SQL Server 2000 data base is used for managing the performance results.

V. EXPERIMENTAL RESULTS

In order to evaluate the performance of our proposed algorithm, we have conducted experiments on a PC (CPU: Intel(R) Core2Duo, 3.16GHz) with 4GByte of main memory running Windows XP. We used 3 data sets. One contains 72 samples (transactions), second one 39 samples and third one 120 samples.

The following shows the results of Association Rule mining using internal characteristics of items.

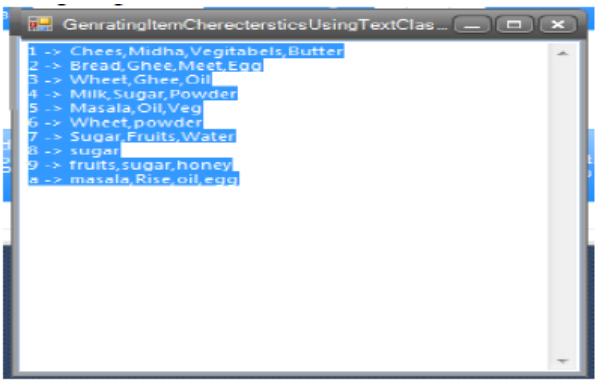


Fig 5.1 Generating internal characteristics of items using text classification.

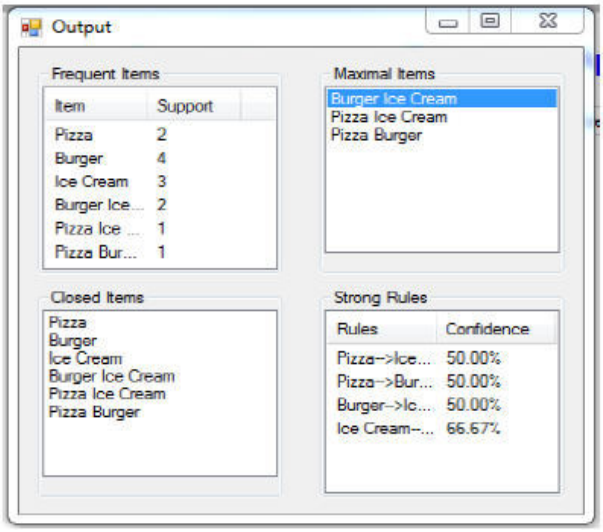


Fig 5.2 Association Rules between items using Apriori

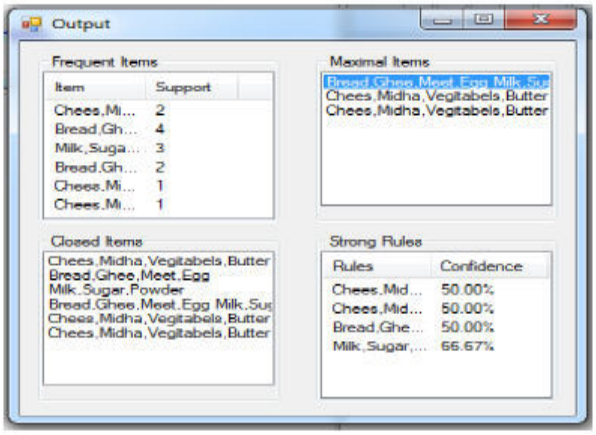


Fig. 5.3 : Association Rules between items characteristics using Apriori

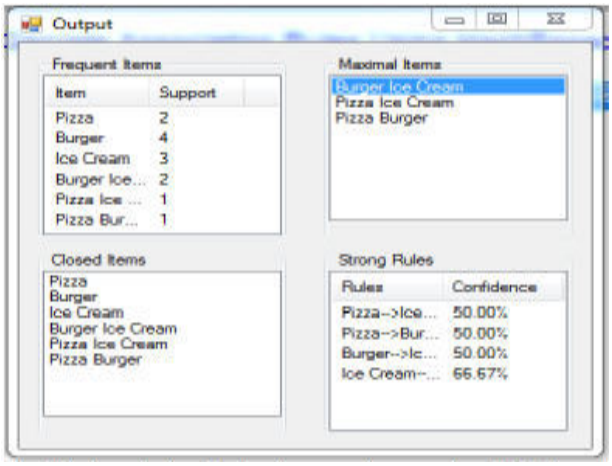


Fig 5.4 Association Rules between items using Hash based Apriori

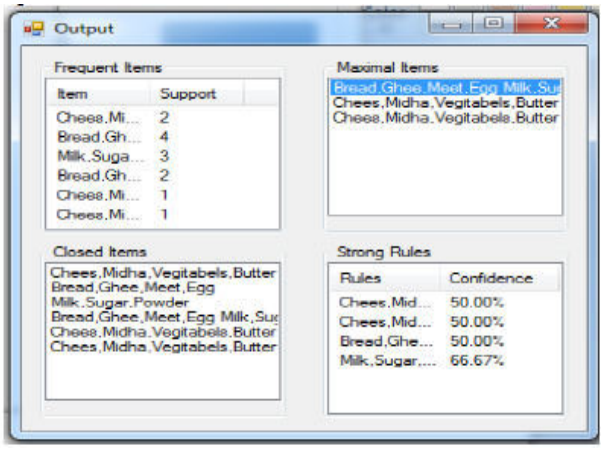


Fig 5.5 Association Rules between items characteristics using Hash based Apriori

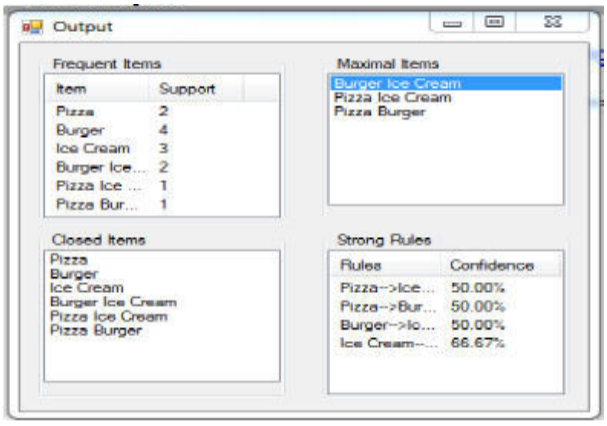


Fig 5.6 : Association Rules between items using FPGrowth

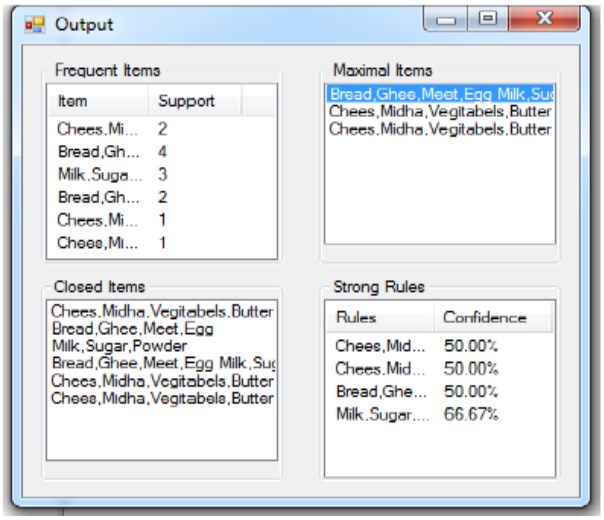


Fig 5.7: Association Rules between items characteristics using FPGrowth

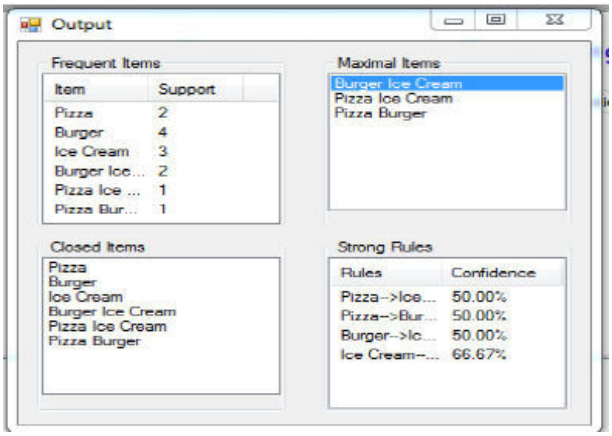


Fig 5.8 : Association Rules between items using FPMax

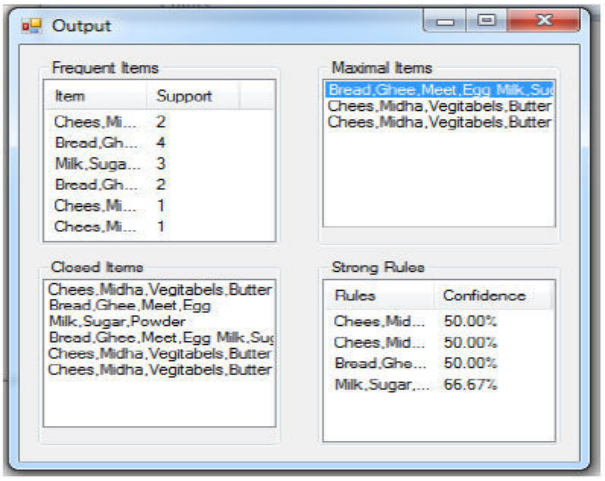


Fig 5.9 : Association Rules between items characteristics using FPMax.

The following diagrams show the comparison of time complexity between various algorithms that generates association rules between items characteristics.

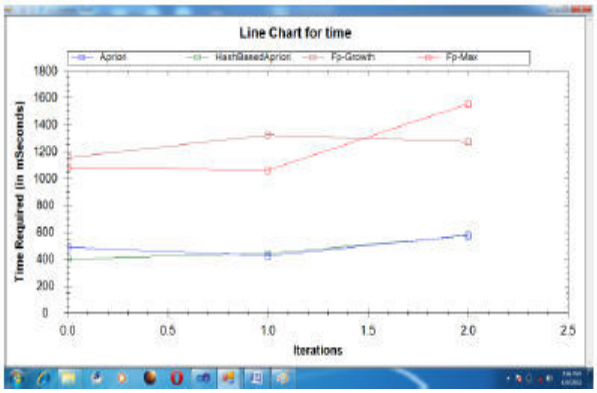


Fig 5.10 : The Line chart Analysis of time complexity.

Fig 5.11 : The Bar chart Analysis of time complexity.

The following diagrams show the comparison of space complexity between various algorithms that generates association rules between items characteristics.

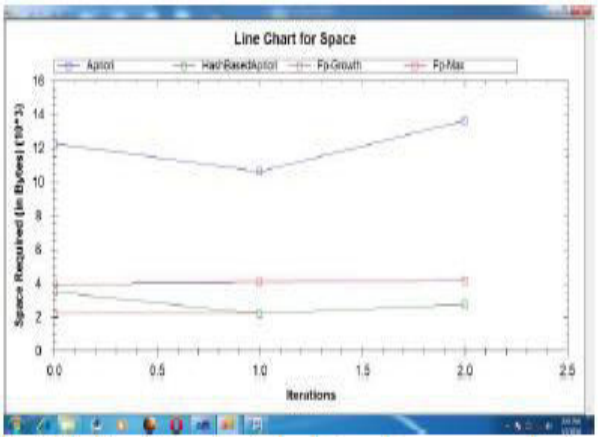


Fig 5.12 : The Line chart Analysis of space complexity.

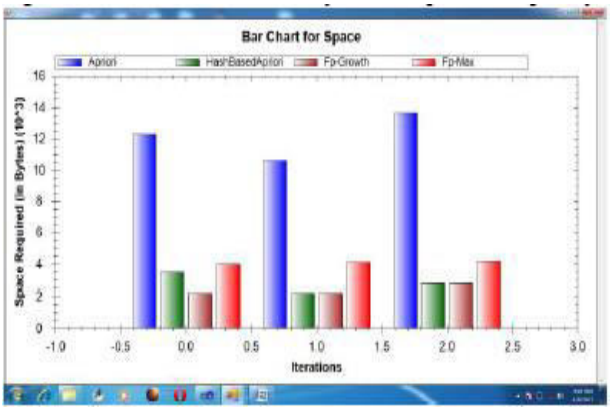


Fig. 5.13 : The Bar chart Analysis of space complexity.

VI. CONCLUSION AND FUTURE WORK

In this paper, we proposed a new approach for Market Basket Analysis by using text classification and Association Rule mining. Text classification extracts the item internal characteristics and association rule mining provides co relation between item characteristics and accordingly the patterns are identified. Our approach provides efficient and accurate frequent patterns which analyzes the performance of customer purchasing behavior.

REFERENCES

[1] J. Han and M. Kamber, "Data mining concepts and techniques(2ed.)," Morgan Kaufmann, pp. 227-275,2006.

[2] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," In Proc. 1993 ACM-SIGMOD Int. Conf. Management of Data(SIGMOD 93), ACM Press, May 1993, pp. 207-216.

[3] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," In proc. 1994 Int. Conf. Very Large Data Bases (VLDB 94), ACM Press, Sept. 1994, pp. 487-499.

[4] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," In Proc. 2000 ACM-SIGMOD Int. Conf. Management of Data(SIGMOD 00), ACM Press, May 2000, pp. 1-12.

[5] H. Zhang, H. Yu, and D Xiong, "HHMM-based Chinese lexical analyzer ICTCLAS," In Proc. Second of SIGHAN Workshop on Chinese Language Processing(ACL 03), ACL press, July 2003, pp.184-187, doi:10.3115/ 1119250.1119280.

[6] H. Zhang and Q. Liu, "Chinese Lexical Analysis Using Hierarchical Hidden Markov Model," In Proc. Second of SIGHAN Workshop on Chinese Language Processing (ACL 03), ACL press, July 2003, pp.63-70.

◈◈◈

Scavang Web Querying for Web People Search

Uma Nashte, Nency Rawal & Bhavin Goradiya

Vidyalankar Institute Of Technology, Sardar Patel Institute Of Tecnology
E-mail : umanashte@gmail.com, nency.patel88@gmail.com, bgoradiya@rediffmail.com

***Abstract* -** Primary interest of this paper is automatic identification of web presence of particular people. The problem appears to have a trivial solution when the name of the person is unique we can just goggle this person's name. However, the problem complexity grows significantly with the level of commonness of the personal name. Indeed, given a common name such as Tom Mitchell we find hundreds of different people called Tom Mitchell. The task of disambiguating and finding the WebPages related to the specific person of interest is left to the user.

In this paper, we propose a Web Appearance Disambiguation (WAD) system to solve the problem by using Tf/Idf similarity and Web querying. The approach is based on extracting named entities from the web pages by using Alchemy API and then use TF/IDF similarity between web pages and querying the web to collecting co-occurrence statistics, which are used as additional similarity measures.

General Terms: *Algorithms,Experimentation,Measurement*

Keywords: *Clustering, Web People Search, WePS, Named Entity Web Co-Occurrences, Social Network Analysis, Web Querying*

I. INTRODUCTION

The rapid growth of the Internet has made the Web a popular place for collecting information. Internet users access billions of web pages online using search engines. Searching for a specific person is one of the most popular search queries.

Given a common name such as "Tom Mitchell" we find hundreds of different people called Tom Mitchell. The main complication is that a web mining system cannot a priori know how ambiguous one or another personal name is. Yet another complication is that the person we are looking for may not have *any* web presence, but his or her namesakes have. And even if the person is well presented in the web, an existence of his of her famous namesake can make the search practically impossible (compare "Julia Roberts" and "Julia Roberts" Professor WKU).

Some additional information about the person is required. When we search for a person in the web, we usually construct a query that provides some kind of summary of the persons activities, e.g. "Tom Mitchell" Professor CMU. However, if our query is too common (such as "Tom Mitchell" Professor), our first hit will be Tom Mitchell the Chicago Professor, and if our query is too narrow (as "Tom Mitchell" "Professor CMU"), we may not find anyone. Since modern search engines are so sensitive to small variations of the same query, it is hard to require an automatic system to construct queries of high quality.

We noticed however that the problem of personal name disambiguation becomes much easier if we are given a few names of people who are known to be related to each other. The goal of a Web People Search (WePS) system is to automatically cluster the WebPages in D such that each cluster corresponds to a namesake.

The proposed solution employs two-phase clustering. First, initial clusters are formed by merging using TF/IDF cosine similarity with a conservative threshold on extracted Named Entities (NEs). This step results in many clusters which are pure but incomplete, that is, they do not contain all of the WebPages of the corresponding namesake. Second, for the pairs of WebPages di and dj that are not merged during the first phase, the algorithm forms additional queries to a web search engine to collect additional similarity evidence in the form of co-occurrence statistics. The returned counts are then transformed to form similarity features for the di, dj pair. We Propose single-link clustering or correlation clustering algorithm to form the final cluster.

II. APPROACH OVERVIEW

In this section we first cover the basic steps of the approach in Section 2.1. We then describe the most important clustering step in more detail in Section 2.2.

2.1 Steps of the Approach

The steps of the overall WePs approach, in the context of middleware architecture, are illustrated in Figure 1.

They include:

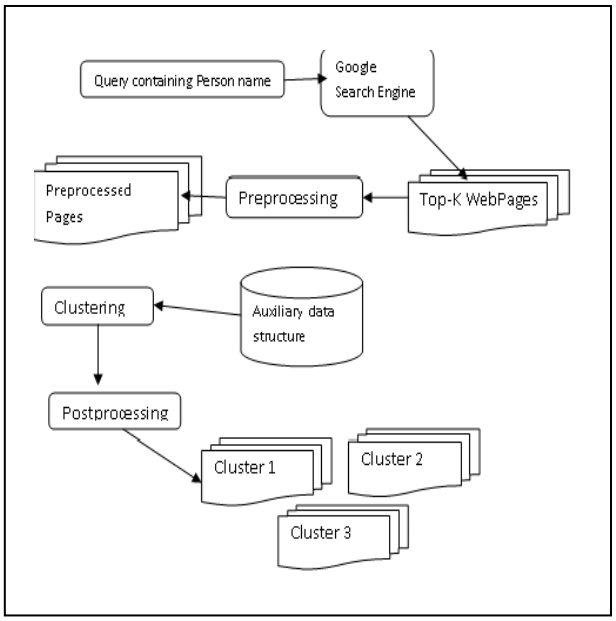


Figure 1: Web Processing Steps

1. User Input. The user issues a query via the input interface.
2. Top-K Retrieval. The system (middleware) sends a query consisting of a person name to a search engine, such as Google!, and retrieves the top-K returned web pages. This is a standard step performed by most of the current WePS systems.
3. Pre-processing. These top-K WebPages are then pre-processed. The main pre-processing steps are:

(a) Extraction. Named Entities, specifically people, locations, organizations are extracted using a third party named entity extraction software (i.e., Alchemy API). Alchemy API is capable of identifying people, companies, organizations, cities, geographic features, and other typed entities within your HTML, text, or web-based content.Some auxiliary data structures are built on this data.

4. Clustering. The top-K WebPages are then clustered. The corresponding algorithms will be explained in more

Detail in Section 2.2. The clustering part consists of the following stages:

(a) TF/IDF Similarity. TF/IDF similarity on NEs only is computed.

(b) Querying the Web. For each pair of WebPages di and dj several co-occurrence queries are formed and issued to a Web search engine.

(c) Feature Generation. The returned co-occurrence counts are then transformed into features.

(d) Making Pair wise Decisions. Using the above features a Skyline-based classifier decides for each pair of WebPages di and dj whether they corefer (positive decision) or not (negative decision).

(e) Applying Clustering. A clustering algorithm, such as single-link clustering or correlation clustering, takes into account the TF/IDF similarity, generated features, and the decisions made by the Skyline-based classifier to output the resulting clustering.

5. Post-processing. The post-processing steps include:

(a) Cluster Sketches are computed.

(b) Cluster Rank is computed based on (a) the context keywords, if present and (b) the original search engine's ordering of the WebPages.

(c) Webpage Rank is computed to determine the relative ordering of WebPages inside each cluster.

6. Visualization. The resulting clusters are presented to the user, which can be interactively explored.

2.2 Clustering

The most important part of the WePS system is clustering. It consists of the following steps.

TF/IDF Similarity. The algorithm computes TF/IDF similarity of the top-k WebPages. The similarity is computed only on the named entities extracted from these WebPages, as general overall text of the WebPages context was found to be too ambiguous, frequently leading to the wrong merge decisions.

The term count in the given document is simply the number of times a given term appears in that document. This count is usually normalized to prevent a bias towards longer documents (which may have a higher term count regardless of the actual importance of that term in the document) to give a measure of the importance of the term *t* within the particular document *d*. Thus we have the term frequency tf (t,d), defined in the simplest case as the occurrence count of a term in a document

The inverse document frequency is a measure of the general importance of the term (obtained by dividing the total number of documents by the number of documents containing the term, and then taking the logarithm of that quotient). $\mathrm{idf}(t) = \log \frac{|D|}{|\{d : t \in d\}|}$
Mathematically the basis of the log function does not

matter and constitutes a constant multiplicative factor towards the overall result. Then

$$\text{tf-idf}(t, d) = \text{tf}(t, d) \times \text{idf}(t)$$

A high weight in tf–idf is reached by a high term frequency (in the given document) and a low document frequency of the term in the whole collection of documents; the weights hence tend to filter out common terms.

Querying the Web.

Unlike many other WePS solutions, including our own [15, 18], the solution we have used for WePS2 does not limit its analysis to the information stored in the Top-k returned web pages. Rather it employs the Web as an external data source to get additional information that could be used to compute similarity of the WebPages. The purpose of using Web queries is to evaluate the degree of interaction of the social networks for two namesakes represented by WebPages di and dj . If there is evidence on the web that two social networks are closely related, then two WebPages are merged into one cluster. The guiding principles in formulating the queries are:

Get-Web-Counts(di, dj)

Let:

N be the queried name

Pi1, Pi2, . . . , Pim be the people NEs extracted from di

Pj1, Pj2, . . . , Pjm be the people NEs extracted from dj

Oi1, Oi2, . . . ,Oim be the org. NEs extracted from di

Oj1,Oj2, . . . , Ojm be the org. NEs extracted from dj

1 Pi (Pi1 OR Pi2 OR · · · OR Pim)

2 Pj (Pj1 OR Pj2 OR · · · OR Pjm)

3 Oi (Oi1 OR Oi2 OR · · · OR Oim)

4 Oj (Oj1 OR Oj2 OR · · · OR Ojm)

5 cij1 GetWebCount (N AND Pi AND Pj)

6 cij2 GetWebCount (Pi AND Pj)

7 cij3 GetWebCount(N AND Pi AND Oj)

8 cij4 GetWebCount(Pi AND Oj)

9 cij5 GetWebCount(N AND Oi AND Pj)

10 cij6 GetWebCount(Oi AND Pj)

11 cij7 GetWebCount(N AND Oi AND Oj)

12 cij8 GetWebCount(Oi AND Oj)

Figure 2: Algorithm for Querying the Web.

Two major types of Web queries are utilized:

1. N AND Ci AND Cj

2. Ci AND Cj.

Here, Ci represents the context for di. It can be either the set of people NEs Pi, or organization NEs Oi. Context Cj is defined similarly for document dj . Since Ci and Cj can have two possible assignments each, this creates 4 context combinations. Given that there are 2 types of queries, this leads to 8 queries in total. For example, assume that the user searches for the WebPages related to “William Cohen”. Suppose that the algorithm extracts 2 namesakes of each type per webpage, that is, m = 2. Assume that webpage di contains names “Jamie Callan”and“Tom Mitchell”, and webpage dj contains names “Andrew McCallum”and“Andrew Ng”. Then the first query will be: “William Cohen” AND (“Jamie Call an” OR “Tom Mitchell”) AND (“Andrew McCallum” OR“AndrewNg”). The web search engine API has a function call that computes the number of WebPages relevant to the query, without actually retrieving those WebPages. Observe that dataset D alone might not have any evidence to merge di and dj . For instance, the TF/IDF similarity between di and dj might be low. Also, among the WebPages in D, names “Jamie Call an” and “Tom Mitchell”might be only mentioned in di, whereas“Andrew McCallum” and“Andrew Ng”only in dj , and otherwise D might not contain any information revealing interactions among these people.

However, querying theWeb allows the algorithm to gain additional information to support the merge. In this case, the counts will be high enough to indicate that the people mentioned in the query are closely related.

Feature Creation. To estimate the degree of overlap of two contexts Ci and Cj for WebPages di and dj for the queried name N we can compute the co-occurrence count $|N \cdot Ci \cdot Cj|$ for query $N \cdot Ci \cdot Cj$. However, it might be difficult to interpret this absolute value without comparing it to certain other values. For instance, if this count value is high, does it mean the contexts overlap significantly and thus di and dj should be merged? Or, is it simply because N is a common name and thus there are lots of WebPages that contains it under many contexts? Or, is it because contexts Ci and Cj are too unspecific, and thus too many WebPages contain them?

The proposed algorithm uses the Dice similarity to get the normalized version of $|N \cdot Ci \cdot Cj|$ count:

In [5] we explain the various advantages of using this specific formula. Making Pair wise Decisions. The web features are then used to make merge or do-not-merge decisions for each pair of pages di and dj . The observation is that the larger the value of the feature the more evidence is there to merge. Thus, if feature f1 dominates feature f2 and f2 has been decided to be classified as a positive merge decision, then logically f1 should also be classified as a positive merge decision, since it has more evidence to merge. Thus, any classifier

that takes this rule into account will essentially learn a classification Skyline, where everything is on and above the skyline should be classified as a merge. One of the key contributions of [5] is a new Skyline-based classifier for deciding which di and dj WebPages should be merged based on the corresponding feature vector. It is a specialized classifier that we have designed specifically for the clustering problem at hand. It learns the Skyline directly from data using supervised learning. Skyline-based classifier gains its advantage due to a variety of functionalities built into it, including:

- It takes into account dominance that is present in the feature space.
- It also fine tunes itself to the quality measure being used.
- It takes into account transitivity of merges: that is, accounts for the fact that two large clusters can be merged by a single merge decision, and, thus, one direct merge decision can lead to multiple indirect ones.

These properties allow it to easily outperform other classification methods (which are generic), such as DTC or SVM.

The approach is discussed in detail in [5].

Applying Clustering. A clustering algorithm takes the information collected above and produces the resulting clustering. The clustering algorithm we used for the WePS2 submission is a single-link hierarchical agglomerative clustering. It first merges all of the WebPages having very high TF/IDF similarity. It then merges all of the pairs WebPages which the skyline based classifier classifies as a "merge".

III. ALGORITHM

Process-Query (Q, k)

1. Get top k web page by using Google API GwebSearchClient().
2. Extract content of each web page by using Alchemy API.
3. Read each web page and analyze whether it contains the queried name N with different middle initials .if so segment web page for this look at html tag<p> before the next full name but after the current name.then segments the document at that point so now your web page becomes K+1.
4. For each web page extract location and organization using alchemy API. Store location and organization name for all pages in one auxiliary database.
5. For each entity in database calculate IDF=log10(D/d)

D=total no of documents

D=no of documents containing word.

6. for each web page for all entity in database calculate TF=(no of times entity occurring in page / total words in page)
7. for each web page for all entity in database calculate Tf-Idf=TF*IDF.
8. Apply cosine similarity formula

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}$$

A1=TF-idf for first entity in database for 1st page

B1=TF-idf for first entity in database for 2nd page

9. for K web pages

   For i=1 to k-1

   For j=i+1 to k;

   S=compute cosine similarity(di,dj);

   Di,dj are web pages

10. if s>t then

Merge In one cluster(di,dj)

// All pairs are "unprocessed" initially.

10for each distinct unmerged unprocessed pair di, dj 2 D do

11. cij Get-Web-Counts (di, dj)

12. fij Convert-To-Features (cij)

13.if Classify-Feature(fij) = merge then

14MergedInOneCluster(di, dj)

15.MarkAsProcessed(di, dj)

16. Visualize-Results-To-User(D)

IV. CONCLUSION AND FUTURE WORK

This paper proposes a Web People Search approach that is based on collecting the co-occurrence information from the Web in order to get a better disambiguation quality.

In this paper we describe our experience of applying our skyline based classification approach proposed in [5] on WEPS-2 dataset. The approach employs the Web as an external data source to collect additional similarity information to better the quality of WePS.

As future work we plan to develop a new solution that would utilize the plethora of other features available in the data, including hyperlinks, emails, phone number, and so on.

REFERENCES

[1] J. G. Javier Artiles and S. Sekine. *Weps 2 evaluation campaign: overview of the web people search clustering task.* In 2nd Web People Search Evaluation Workshop (WePS 2009), 18th WWW Conference, April 2009.

[2] D. V. Kalashnikov, Z. Chen, S. Mehrotra, and R. Nuray. *Web people search via connection analysis.*

IEEE Transactions on Knowledge and Data Engineering (IEEE TKDE), 20(11), Nov. 2008.

[3] D. V. Kalashnikov, Z. Chen, R. Nuray-Turan, S. Mehrotra, and Z. Zhang. *WEST: Modern technologies for Web People Search.* In Proc. of the 25th IEEE Int˝Sl Conference on Data Engineering (IEEE ICDE 2009), Shanghai, China, March 29 - April 4 2009. demo publication.

[4] D. V. Kalashnikov, S. Mehrotra, S. Chen, R. Nuray, and N. Ashish. *Disambiguation algorithm for people search on the web*. In Proc. of the IEEE

[5] D. V. Kalashnikov, R. Nuray-Turan, and S. Mehrotra. *Towards breaking the quality curse. A web-querying approach to Web People Search.* In Proc. of Annual International ACM SIGIR Conference, Singapore, July 20–24 2008. April 9–12 2007.

◈◈◈