IJACSA
WHERE WISDOM SHARES
International Journal of Advanced Computer Science and Applications

Volume 3 Issue 10

October 2012

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

SAI

www.ijacsa.thesai.org

INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS

THE SCIENCE AND INFORMATION ORGANIZATION

www.thesai.org | info@thesai.org

Editorial Preface

From the Desk of Managing Editor…

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

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

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

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

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

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

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

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

Thank you for Sharing Wisdom!

Managing Editor
IJACSA
Volume 3 Issue 10 October 2012
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

Reviewer Board Members

CONTENTS

Challenges of Future R&D in Mobile Communications

Dr. Anwar M. Mousa
University of Palestine-Gaza- Palestine
Tel +970 599 609918 Fax +972 8 28 80006

***Abstract*— This paper provides a survey for the main challenges of future research and development (R&D) for next generation mobile networks (NGNs). It addresses software and hardware re-configurability with focus on reconfigurable coupling and interworking amongst heterogeneous wireless access networks. It also explores the promising technologies for NGNs such as nanotechnology, cloud computing and texting by thinking. The paper then highlights the challenging research areas for enhancing network performance and affordability beginning with All IP network and security issues, Vehicular Ad Hoc Networks (VANET) and the necessity of high data rates at cell edges. Investigating the viability of cooperative networks and one unified global standard for NGNs, the paper analyzes macro-diversity and advanced multi-cell coordination for mitigating inter-cell interference. Direct device-to-device communication, global coverage using satellites and high-altitude platforms are presented as possible evolution paths for space roaming and extending NGN' radio-access. Finally, the paper inspects the most developing applications for NGNs such as World-Wide Wireless Web (WWWW), machine-type communication, wireless sensor networks (WSNs), e-healthcare systems and wearable devices with AI capabilities.**

Keywords- R&D; NGNs; re-configurability; VANET; nanotechnology; cloud computing; HAP; WWWW; WSNs; e-healthcare; wearable devices.

I. INTRODUCTION

The key concepts of next generation mobile networks (NGN) suggested in research papers are:

- Full interworking (interoperability) of heterogeneous systems.
- Up to 1 Gbit/s of broadband services, delivered by standard/system, for users in any environment [1].
- Real wireless world with no limit in access and zone.
- Wearable devices with AI capabilities.
- IPv6, where the assignment of the visiting care of mobile-IP address is done according to location and connected network [2].
- One unified global standard.
- Cognitive Radio (CR) based technology [3].
- High-altitude platforms station systems
- Wireless sensor systems for surveillance and environmental sensing.

For the time being, the 4-Generation (4G) systems are still under intensive research and development efforts; there are numerous research groups and development projects worldwide aiming at widespread 4G wireless and cellular deployments all over the world. Surely, researches will continue progressing in the future for beyond 4-G systems. This paper highlights main areas of research and development (R&D) challenges that face NGN.

The paper is organized as follows: Section II discusses in details re-configurability in mobile communications with focus on re-configurability enablers and reconfigurable interworking amongst heterogeneous wireless access networks. Reconfigurable (adaptive) coupling of heterogeneous networks is proposed as new coupling method that adaptively changes coupling level from open, loose to tight and even very tight according to networks' load status and delay constraints. Hardware and software re-configurability with different reconfiguration and downloading modes are also presented in section II. Section III explores the promising technologies for NGNs such as nanotechnology, cloud computing and texting by thinking. Network energy efficiency with reduced energy consumption necessary for both mobile terminal and radio access network is thoroughly illustrated. Section IV highlights the challenging research areas for enhancing network performance and affordability beginning with All IP network and security issues, Vehicular Ad Hoc Networks (VANET) and the necessity to high data rates at cell edges. It investigates the viability of cooperative networks and one unified global standard for NGNs. Macro-diversity and advanced multi-cell coordination for mitigating interference are also analyzed. Finally direct device-to-device communication and global coverage using satellites, and high-altitude platforms are presented at the end of section IV. Section V focuses on the most developing applications for NGNs such as World-wide wireless web (WWWW), machine-type communication, wireless sensor networks (WSNs), E-Healthcare Systems and wearable devices with AI capabilities. Section VI concludes the paper.

II. RE-CONFIGURABILITY IN MOBILE COMMUNICATIONS

Re-configurability in mobile communication systems makes use of the abilities of reconfigurable terminal and networks for self-adaptation to a dynamically-changing environment aiming at improving service offering and spectrum utilization. To meet the rapidly increasing demand for mobile services, with the limited radio spectrum, more flexible ways to share radio spectrum among multiple services and radio networks are required. Re-configurability is based

on Cognitive Radio (CR) which in turn is considered as an evolution of Software-Defined Radio (SDR) [4]. A CR is the final point of SDR platform evolution: a fully software and/or hardware reconfigurable radio that changes its communication functions depending on network and/or user demands. A CR is an SDR that additionally senses and tracks changes in its environment and then reacts upon its findings. The implementation of re-configurability in mobile systems spans from reconfigurable interworking (interoperability) among heterogeneous wireless access networks to reconfigurable (Adaptive) coupling of different networks. Re-configurability could be in both software and hardware. The reconfigurable software radio concept has been an active research and development topic for many years. With the power of SDR, network is dynamically reconfigurable; the programs running on the reconfigurable processing elements as well as the communication links between the processing elements are configured at run-time. Different downloading modes in software reconfiguration exist each having its own advantages and disadvantages; mission-oriented re-configurability and self- re-configurability. Hardware re-configurability is based on the evolution of semiconductor technology and mainly performed by operators; adding additional equipments to increase network capacity at a specific time. More illustration in the following subsections:

A. Re-configurability enablers

1) Software Defined Radio

Software Defined Radio (SDR) [5] benefits from today's high processing power to develop multi-mode (multiband and multi-standard) base stations and terminals. In future the terminals will adapt the air interface to the available radio access technology, however at present this is done by the infrastructure. Several infrastructure gains are expected from SDR. For example, to increase network capacity at a specific time (e.g. during festivals or sport events), an operator can reconfigure its network by adding several modems at a given Base Transceiver Station (BTS). SDR makes this reconfiguration easy. For the expected next generation systems, SDR will become an enabler for terminal and network re-configurability through software download. Figure 1 shows a general reconfigurable SDR transceiver in both terminal and base stations. It can be reconfigured via a control bus supplying the processing units with the parameters downloaded from remote re-configuration database via a predefined download channel. For manufacturer, this can be a powerful aid for providing multi-mode equipment with reduced development effort and costs.

Refer Figure 1 in the Appendix: General SDR transceiver

2) Cognitive Radio

Cognitive radio (CR) technology allows different radio access technologies to share the same spectrum efficiently. This is done by adaptively finding unused spectrum and reconfiguring (adapting) the transmission scheme to the requirements of the technologies currently sharing the spectrum. The applicability of CR to cellular communication is a relatively new area and further studies are required to assess the feasibility and impact of such usage. CR technologies includes the ability of equipments to determine their location, sense spectrum used by neighboring equipments, change their frequencies, adjust output power, and even alter transmission parameters. It is a transceiver that is able to understand (think) and react to its operating environment. Thus CR concerns both mobile terminal and networks which are computationally intelligent about radio resources and related communications. The Radio is aware about changes in its environment and responds to these changes by adapting operating characteristics in some way to improve its performance.

A CR technology for spectrum re-uses and network overlay, named Vandermonde-subspace frequency division multiplexing (VFDM) [6] is newly developed. It is a modulation scheme to allow the co-existence of macro-cells and cognitive radio small-cells in mobile network. VFDM, a technique for interference cancellation in overlay networks, allows a secondary network to operate simultaneously with a primary network on the same frequency band. It can be applied to block transmission systems with a guard time over frequency selective channels. It achieves zero interference towards the primary system by employing a specific pre-coder that aligns the data to the null space of the interfering channel from the secondary to the primary system.

B. Reconfigurable Interworking amongst heterogeneous wireless access networks

A seamless interworking amongst heterogeneous network represents the corner stone for the success of next generation systems with different evolving access technologies. A novel solution that ensures interworking between several types of wireless access network is given by the IEEE 802.21 standard [7]. The IEEE 802.21 is focused on handover implementation between different wireless networks in heterogeneous environments regardless of the type of medium. The standard names this type of vertical handover as Media Independent Handover (MIH). The goal of IEEE 802.21 is to facilitate the mobile nodes' usage by providing uninterrupted handover in heterogeneous networks. The heart of the 802.21 framework is the Media Independent Handover Function (MIHF) which is implemented in every IEEE 802.21 compatible device (in either hardware or software). This function is responsible for communication with different terminals, networks and remote MIHFs. It provides abstract services to the higher layers using a unified interface (L2.5 functionalities). Although the IEEE 802.21 standard is still in its formative stages, it may be the key enabler for seamless vertical handover and transparent roaming in heterogeneous networks. Hence, this standard will make a major contribution towards the reconfigurable interworking aspect of next generations wireless and cellular communications systems. The IST Ambient Networks projects [8], uses a single multi-radio system with powerful multi-radio resource management mechanisms and generic link layer functionality to integrate the envisioned future heterogeneous environment. This eases the provision of the "Always Best Connected" [9] approach paving the way towards interworking heterogeneous networks. Besides, the IST MAGNET project [10] and IST MAGNET Beyond project [11] provide a solution of some technological issues concerning interworking between different network interconnection schemes.

The reconfigurable interworking can be implemented at the network level, the user level or both, bringing benefits from both the network providers' perspective and the users' perspective. It also contributes to the robustness of the provisioning of users' requested services allowing user seamless and transparent service management [12]. At the network level, the reconfigurable interworking allows network providers to select, with minimal investments, between alternative wireless access networks. The selection could be made based on several criteria such as:

- Comparison between the availability of radio access resources and service requirements (users' QoS requirements, channel state, outage probability, vertical handover probability…etc).
- Efficient load sharing between different coexisting wireless networks.
- Efficient spectrum sharing.
- Congestion control.

Thus, the changes in the network resource availability due to network instantaneous saturation can be bypassed by terminals and network components that are dynamically reconfigured (adapted) to the new situation. At the user level, the reconfigurable interworking of the heterogeneous systems leads to more efficient end-to-end connectivity and service delivery in heterogeneous environments. It also provides easier worldwide roaming and dynamic adaptation to regional contexts. The users' equipments reconfigure based on:

- Availability of spectrum and radio access resources
- Service cost Minimization when multiple technologies are available.
- Anticipation of user contexts and preferences.

C. *Reconfigurable (Adaptive) Coupling of Heterogeneous Networks.*

Different types of coupling of heterogeneous networks and therefore different integration approaches can be classified as follows:

- Open coupling

With open coupling, no effective integration exists between two or more radio access technologies with only a billing system being shared between them. Separate authentication procedures are used for each access networks and no vertical handovers take place. The only interaction happens between the billing management systems of each network technology, but no interaction exists between the control procedures related to the QoS and mobility management.

- Loose coupling

In loose coupling the operator is still able to use the existing subscriber database, allowing centralized billing and maintenance for different technologies. Besides, there is an interaction between the control planes of each operator regarding the authentication procedure. Hence, the user enjoys a unique subscription if the network provider is the same for both networks. The main disadvantage of loose coupling is that vertical handover is not seamless; during the handover between the two Radio Access Technologies (RATs), the service in progress is dropped. Besides, the handover delay is significant as mentioned above.

- Tight coupling

The advantage of tight coupling is that it permits seamless handover between radio access technologies to take place. However it requires the generic RAT networks to support necessary interfaces with other coupled technologies.

- Very tight coupling

In the very tight coupling, the interworking is provided within the radio access network. Due to a new interface definition between different RATs, the networks are able to perform a seamless inter-network handover. IST-TRUST project [13] focused a great effort on very tight coupling between networks and allowed for very fast vertical handover. Adaptive coupling is a new proposed mechanism that adaptively changes coupling level from open, loose to tight and even very tight according to networks' load status and delay constraints. On the one hand, if the integration between different technologies is tight, the provisioning of the service is more efficient and network selection as well as the vertical handover process is faster. However, a high level of integration requires considerable effort in the definition of interfaces and mechanisms able to support the necessary exchange of data and signaling between different radio access networks. Moreover, tight coupling suffers from potential of load congestion when one network full load is immersed on the other. On the other hand, if the integration between different technologies is loose, the delay of handover process is significant. On the positive side, loose coupling allows for the flexibility and independence of implementing individually different mechanisms within each network. Besides, it eases the gradual deployment of one network with no or little modification on the other network(s).

D. *Software Re-configurability*

Projects in literature dealing with software reconfigurability have focused on the technology for providing multi-mode software reconfigurable terminals. IST TRUST project [13] identified the frameworks and systems that are needed to support software reconfigurable radios from the users' perspective by examining user requirements. Realizing the user potential of reconfigurable radio systems, the project provided network connectivity and services when and where they are required. Hence, simultaneous connections to different RATs became possible. The IST MOBIVAS project [14] enabled flexible provisioning of value-added services in mobile communication networks by developing architectural approaches and prototypical implementations of integrated software platforms. The IST PACWOMAN project [15] enabled the design of low-cost and low-power reconfigurable user terminal by performing the necessary R&D on all OSI layers. Software reconfigurable transceivers, whether at the user terminal or at the base station, differ from a conventional transceiver only by the fact that they scan the available networks and, after downloading the appropriate software(s) reconfigure themselves as per the selected network features.

Different reconfiguration and downloading modes exist as follows:

1) Reconfiguration database download

Software modules downloaded, via Internet, from a remote reconfiguration database using cloud computing. A transceiver can be reconfigured via a control bus supplying the processing units with the parameters which describe the desired standard. Such a reconfiguration guarantees that the transmission can be changed instantaneously if necessary; in case of inter-standard handover.

2) Re-configurability using a dedicated channel

A dedicated channel bridges between user terminal and base station facilitating error-free downloading of the requested software module. However this approach suffers from the varying bandwidth requirements over narrow-band channels and also of the cost of allocating a dedicated physical channel for downloading.

3) Re-configurability using advanced SIM Cards

Downloading and reconfiguring of the appropriate services on the terminals takes place upon insertion of smart cards including prior stored information regarding the reconfigurable software modules. However this approach is mostly suitable for stand-alone terminals.

4) Re-configurability using predefined broadcasting pilot

This approach is based on a predefined broadcasting global pilot and download channel (GPDCH) which is constantly monitored by multimode user terminals. On detection of any available network, the terminals decide whether to switch off or not. This approach suffers from long downloading timedue to slow speed of GPDCH.

5) Mission-Oriented Re-configurability

Reconfiguration to meet a given set of mission requirements is called mission oriented reconfiguration. Typical mission requirements might include operation within buildings, over long distances and while moving at high speed. Mission-oriented reconfiguration usually selects a set of radio software modules from a library of modules and connecting them into an operational radio.

6) Self- Re-configurability

A transceiver normally consists of several modules; a radio frequency front-end, a digital signal processor, and a control processor. In this approach, each module is a self-describing and the radio automatically reconfigures itself for operation from the available modules.

E. Hardware Re-configurability

Hardware reconfigurability cannot be separated from software reconfigurability. However the reconfigurable hardware is based on the evolution of semiconductor technology. More and more transistors can be integrated on a single chip which makes it possible to build large systems on a chip level, System-on-Chip (SoC) [16]. Hardware reconfigurability is made easy thanks to various processing elements such as General Purpose Processors (GPPs), Field Programmable Gate Arrays (FPGAs), Application Specific Integrated Circuits (ASICs) and Domain Specific Reconfigurable Hardware (DSRH) modules. Different processing elements are used for different purposes. The GPPs are programmable to perform different computational tasks. FPGAs which are reconfigurable by nature, are efficient at performing bit-level operations but not for word level DSP operations. The ASICs are optimized for power and cost but cannot be reconfigured to adapt to new applications. The DSRH is a relatively new type of processing element, where the configurable hardware is adapted towards a specific application domain.

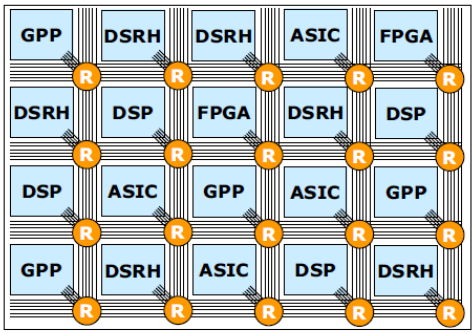


Figure 2: An example of a heterogeneous SoC. *Source: from [16]*

III. PROMISING TECHNOLOGIES

A. Nanotechnology

Nanotechnology is derived from nanometer scale; between 0.1 and 100nm. It uses nano-science to control specific processes. The field is also known as molecular nanotechnology (MNT) where MNT deals with control of the structure of matter based on atom-by-atom and molecule by molecule engineering. Nanotechnology is considered as the coming industrial revolution, and the communications industry will be radically transformed by it in a few years. Current technologies cannot resolve challenges such as more memory and computing power required to accommodate accelerating intelligence, more accurate sensing and increased data rates. Nanotechnology could provide effective solutions for more powerful computing, nano-sensing, memory enlargement, and human-machine interaction [17]. Nanotechnology will have considerable impacts on both mobile terminal as well as core network as follows:

- The mobile terminal has become more than a communication device in modern world; computation and communication are ready to serve the user in an intelligent way. Mobile terminals besides the intelligence embedded in human environments will enable ubiquitous sensing, computing, and communication. The Impact of nanotechnology over mobile terminals enables them to act as intelligent nano-enabled sensors. Nano-enabled sensors have applications in many industries, among them transportation, communications, medicine and safety.

- The core network requires high speed and a reliable capacity to manipulate and interoperate the increasing number of heterogeneous access technologies. At present, nanotechnologies are used in Digital Signal Processing (DSP) Fabrication. Much more new perceptions are introduced in DSP designing which increases the overall system speed & capacity.

Nanotechnology enables higher transfer rate of data within and between devices since a major limitation in transfer rate is the use of electrical wiring and contacts. Technologies under development that will lay the foundation for faster data transfer, more mobile processing power and larger data storage are:

- Nano-materials with novel optical, electrical, and magnetic properties.
- Faster and smaller non-silicon-based chipsets, memory, and processors.
- Computers based on Quantum Computing.
- Advanced manufacturing and microscopy systems.
- Smaller and faster optical switches.
- Nano-scale MEMS: micro-electro-mechanical systems.

B. Network Energy Efficiency

Reduced energy consumption is necessary for both mobile terminal and radio access network as follows:

1) Mobile Terminal

Low energy consumption for mobile terminals has been a critical concern since the emergence of hand-held terminals roughly 25 years ago. The driving force has been the reduction in battery size and improved battery time. A new nano-technology called Power Felt is an innovative way to recharge battery by converting body heat into electric current. Hence the battery can be recharged at any time. Felt the power itself, made of carbon nano-tubes, is very small and locked in flexible plastic fibers that resemble fabric. This technology uses the difference between room temperature and the temperature of the body to create energy. Generally, thermo-electric is not yet fully developed technology to produce energy. Besides, the high cost of such technology makes it unfeasible to use in consumer products. Other promising method for recharging the mobile terminal is to be self powered; the phone derives its energy/power from the sun where it shines.

2) Radio-Access Network

Today, a close attention is paid to reduced energy consumption also in the radio-access network. This is due to the resultant global warming and increased energy consumption in the information and communication technology (ICT) industries. Energy efficiency has, hence, become a key performance metric to evaluate the performance of a communication network. To prevent our planet from global warming and effectively increase energy efficiency, it is expected that future wireless networks will use eco-friendly green energy. In some rural areas, it may not even be possible to connect the base station to the electrical grid. Instead of the diesel generators commonly used today, solar panels could be used as power source for low consumption. Moreover, the cost of energy is a considerable part of the overall operational cost for the operator. Nevertheless, the future evolution of cellular systems should further minimize unneeded transmission of signals by designing energy-aware routing algorithms with sustainable power models.

C. Cloud computing

Cloud computing [18] is a technology that uses the internet and central remote server to maintain data and applications. It is a model for enabling ubiquitous, on-demand access to a shared pool of reconfigurable computing resources such as servers, storage, applications, and services. In next generations networks this central remote server could be a content provider. Cloud computing allows consumers and business to use applications without installation and access their personal files at any computer with internet access. Cloud computing promotes service development where operators can enter the cloud computing market and create new value-added services by integrating industry content and applications in the digital supermarket model.

Figure 3: Cloud Computing, Ubiquitous Access. *Source: from [19]*

D. Texting by thinking

In future mobile generation, phones may become available with the option of texting by thought power alone. Even though one may think this is imaginative, it could be achieved by a sensor-mounted headset worn by the user. The device contains brain-machine interface technology which analyses brain waves, converts them into digital signals and displays the resulting letters on-screen. This allows completely hands-free texting, effectively creating a form of electronic telepathy. Even though the process is rather slow at this stage, nevertheless advances in the coming years will enable smooth and fast interactions, revolutionizing the world of communication.

IV. Enhanced Network Performance And Affordability

This section highlights the challenging research areas for enhancing network performance and affordability beginning with All IP network and security issues, Vehicular Ad Hoc Networks (VANET) and the necessity to high data rates at cell edges. It investigates the viability of cooperative networks and one unified global standard for NGNs. Macro-diversity and

advanced multi-cell coordination for mitigating interference are also analyzed. Finally direct device-to-device communication and global coverage using satellites, and high-altitude platforms are presented.

Figure 4: Texting by thinking, revolutionizing the world of communication.

A. All IP Network and Security Issues

The All-IP Network (AIPN) is an evolution of the 3GPP system to meet the increasing demands of the mobile telecommunications market. It is a common platform valid for all sorts of radio access technologies. Primarily focused upon enhancements of packet switched technology, AIPN provides a continued evolution and optimization of the system concept in order to provide competitive edge in terms of both performance and cost. The key benefits of AIPN architecture includes a variety of different access systems' provision, lower costs, universal seamless access, increased user-satisfaction and reduced system latency. But with the advantages of IP come some dangers: as data flow more freely and the internet is open not only to developers but also to all manner of criminals and viruses, developers and operators face new security challenges which should be solved properly. Typical issues for network security and privacy could be how to achieve data confidentiality, integrity and authenticity. The current solutions are the use of cryptographic primitives (such as encryption, signature, and key agreement) or the use of anonymity techniques (such as multiple pseudonyms). However, security and privacy issues considered in a specific network cannot be well solved by directly adopting those techniques. The solutions must take into account the typical characteristics of the network.

Refer Figure 5 in Appendix: Core IP Network, different access systems' provision.

B. VANET Communications

Vehicular Ad Hoc Networks (VANET) have been envisioned to play an important role in the future wireless service market for safety communications, information and entertainment applications [20]. Examples of safety-oriented applications for VANET are the notifications of emergency situations, such as car accidents or bad weather conditions. Applications in VANET can be implemented by using vehicle-to-infrastructure (V2I) communications such as the downloading of music and video files, or vehicle-to-vehicle (V2V) communications, such as in distributed games played among passengers in neighboring vehicles. Challenges of future research in VANET may focus on IP mobility and IPv6 deployment in VANET, capacity and mobility connectivity, mobility models for vehicular networks and data dissemination.

C. High data rates available at cell edges

A major challenge of future generation systems is to make the high bit rates available in a larger portion of the cell, especially to users in cell edges and in an exposed position in between several base stations. In current research, this issue is addressed by cellular repeaters and macro-diversity techniques, also known as Group Cooperative Relay (GCR), as well as by Beam Division Multiple Access (BDMA). In spite of the fact that the current OFDMA is appropriate for next generation systems' radio access, BDMA and GCR are also gaining significant attention as effective radio access techniques.

D. Cooperative Networks

In the future, services should be delivered seamlessly anytime, anywhere over heterogeneous networks. As an effective solution, cooperative networking has received significant attention recently. In cooperative networking, individual network entities operate in a coordinated way to implement network goals. It spans both the wireless access segment and the backbone information transport and delivery segment.

E. One unified global standard.

The WiMAX Forum [21] had previously promoted IEEE 802.16m [22] as their solution for a true 4G standards while 3GPP had promoted a similar, but incompatible, 4G standard called LTE Advanced [23]. The 4G Forum would create a single International Telecommunications Union (ITU) standard, world-wide. The incompatible Mobile WiMAX and LTE standards would then evolve into one single interoperable standard called IMT-Advanced [24]. A single IMT-Advanced unified standard will help address bandwidth shortages. Currently, incompatible 4G standards need to be separated from each other to avoid interference. In addition, incompatible systems increase cost and decrease performance. Until the single IMT-Advanced standard becomes reality, the currently incompatible WiMAX and LTE standards are under heavy research to evolve into a single interoperable NGN standard with 100 Mbps (mobile) and 1 Gbps (fixed) speeds, world-wide.

F. Macro Space-Diversity and Advanced Multi-Cell Coordination

Macro space-diversity refers to the transmission and/or the reception at multiple geographically separated sites. The aim is to enhance signal quality and reduce interference. Related to this field is the Coordination of Multi-Point transmission and reception (CoMP) [23]. The umbrella of CoMP, comprises many different coordination schemes of very different characteristics: ranging from dynamic inter-cell scheduling coordination (DISC) to joint transmission/reception at multiple sites. In DISC case, CoMP can nearly be seen as an extension of the inter-cell interference coordination. In joint transmission, data is transmitted to a mobile terminal jointly from several sites, thereby not only reducing the interference but also increasing the received power. Similarly, in joint reception the signals received at multiple sites are jointly

processed for enhanced reception performance. Maximum-ratio combining and interference-rejection combining are efficient schemes used to combine the uplink transmission received at multiple points. This is, in many respects, similar to softer handover used within a site, for example in WCDMA/HSPA [25] based systems, but extended to multiple sites. Challenges for future research, regarding joint reception and transmission concerns lower latency in the communication between the network node involved in the joint processing and the different antennas involved in the reception/transmission.

G. Direct Device-to-Device Communication

Direct Device-to-Device (D2D) communication is considered as one possible evolution path to extend NGN' radio-access technology. That is, making it possible for two mobile terminals to communicate directly with each other without going via the network. This has been studied in academia for some years, for example as part of the European research project WINNER [26].Emergency situations and national security and public safety are possible scenarios where the direct D2D communication could be of interest. Moreover, situations where no network infrastructure is available and situations where network infrastructure is available but when communication directly between the terminals could be more efficient are other scenarios for efficient applications of direct D2D communications. Furthermore,, an attractive application for D2D communication is the possibility to use NGNs as a radio-access technology at home. When outside home, the devices can communicate with a cellular network, but while at home they can be connected, even without a SIM card, to the local network covering the home. However, this area is still in its beginning and scenarios where D2D communication offers benefits and how to interact with infrastructure-based communications pose challenges for future research.

H. Global Coverage : Satellites, and High-altitude platforms

Very high data rates and more service applications could be allowed by the integration of satellites and terrestrial mobile communications. The global coverage systems, achieved by this integration, have been developed by four courtiers. The global position system (GPS) has been developed by USA, the COMPASS system by China, the GALILEO system by EU and the GLONASS by Russia [27]. These independent systems are difficult for space roaming hence a challenge of NGNs is to unite the four systems to get space roaming. At lower altitudes than satellites, high-altitude platforms (HAP) are receiving more and more attention for providing coverage to a large area while staying tens of kilometers above in the air for long periods of time [28]. HAP is a quasi-stationary aircraft; it is specially designed to operate at a very high altitude (17–22 km) and is able to stay there for days. However, the new generations of HAPs will expand this period to several years! This goal is a challenges for future research. Instead of powering current HAPs by batteries or engines, which need recharging or refueling, solar energy is one of best options currently being used for under trial HAPs. HAP can be a manned or unmanned and is capable to cover a small region much more effectively than satellites because it operates at much lower altitudes. Lower altitude also means much lower power consumption and smaller delay compared to satellites. Furthermore, deploying a satellite drains significant time and monetary resources, in terms of development and launch. HAPs cost much lower than satellite and are rapidly deployable. The applications of HAPs cover the following areas:

- *High-speed wireless communications,* Research on HAPs is being actively carried largely in Europe to deliver high speed connectivity to users, over areas of up to 400 km. It has gained significant interest because HAPs will be able to deliver data rates similar to a broadband wireless access network (such as WiMAX) while providing a coverage area similar to that of a satellite.
- *Surveillance, security and real-time monitoring,* one example of a HAP used for surveillance and security is RQ-4 Global Hawk UAV [29]. It possesses a highly sophisticated sensor system and is able to deliver digital sensor data in real-time to a ground station. Powered by a turbofan engine, it can stay in the air for continuous 36 hours. another future use for HAP, currently being under continuous development, is monitoring of a particular region for activities such as flood detection, seismic monitoring, remote sensing and disaster management.
- *Weather and environment monitoring,* the most common use of HAPs is for measuring environmental changes or to keeping track of weather. Recently, NASA in partnership with The National Oceanic and Atmospheric Administration (NOAA) has started using the aforementioned Global Hawk UAV to study Earth's Atmosphere.

A special form of HAPs is High Altitude Stratospheric Platforms (HASP), which are low cost platforms able to fly at altitudes ranging between 17 and 20 km for weeks to months. They are attractive for the provision of future personal communication services. The derivation of a sophisticated channel model for the communication link between the platform and terrestrial mobile users or stations constitutes a challenge for future research.

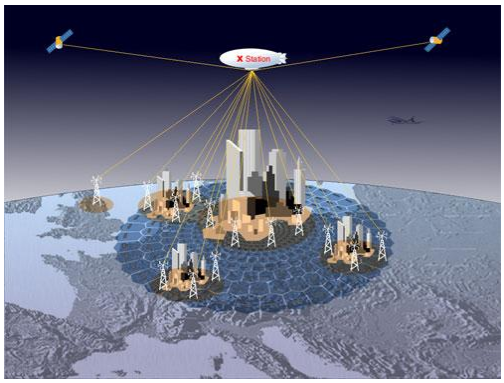


Figure 6: HAP (balloons) linked with various terrestrial and satellite networks.

V. DEVELOPING APPLICATIONS

This section focuses on the most developing applications for NGNs such as World-wide wireless web (WWWW), machine-type communication, wireless sensor networks (WSNs), wearable devices with AI capabilities and E-Healthcare systems.

A. World-Wide Wireless Web (WWWW)

WWWW are comprehensive wireless-based web applications that include multimedia capability that are beyond 4G speeds. The wireless Web refers to use of the World Wide Web through a wireless terminal, such as a cellular telephone or personal digital assistant (PDA). WWWW is expected to provide anytime and anywhere connection to e-mail, mobile banking, instant messaging, weather, travel information, and other services. In general, sites aiming to accommodate wireless users must provide services in a format displayable on typically small wireless terminals. WWWW is widely considered as a cornerstone for NGNs.

Figure 7: WWWW, comprehensive wireless-based web applications.

B. Machine-Type Communication

Various types of connectivity has become a realistic option for machine-type communication due to the increased availability of mobile broadband nowadays. Applications of machine-type communication, currently handled by 4G systems, spans a wide range of applications: from massive deployment of low-cost battery-powered sensors to remote-controlled utility meters, to surveillance cameras. However, other unlimited applications pose challenges in terms of a vast amount of devices connecting to the network. Handling such a large number of devices, improvements in the area of connection setup and power efficient handling of control signaling in the radio-access network may be of great interest for research and development of NGN.

C. Sensor Networks

Wireless sensor networks (WSNs) are designed by grouping mini and/or micro-sized sensors that are equipped with micro-processors, necessary memory and radio transceivers. Sensors are powered by low-energy batteries. WSNs can measure distance, vehicle direction and speed, wind direction and speed, humidity, temperature, chemicals, light, vibrations, seismic data, acoustic data, strain, torque, load, pressure, and other things. WSNs have emerged as a promising solution to a wide range of applications that would redefine the way in which man lives and works. The emergence of nano sensors and nano-enabled sensors have added more applications in many industries, among them transportation, communications, medicine, safety, and security. However, more elaborations of nano sensors for widespread and commercial usage constitute challenges for future research.

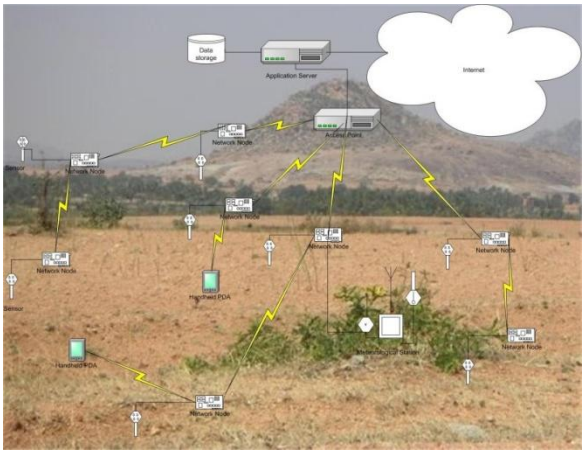


Figure 8: WSNs, a promising solution to a wide range of applications.

D. Wearable devices with AI capabilities.

Wearable devices are mini electronic devices, worn by the human being under or on top of clothing and it could be even inserted under skin, Figure 9. They are related to the field of ubiquitous computing and the development of wearable computers. With ubiquitous computing upgraded with AI capability, wearable devices enter our everyday life. They are able to perform multi-tasks without needing to stop what one is doing! They can therefore be an extension of man's mind and/or body. The future development of the wearable devices goes in line with mobile computing, ambient intelligence and ubiquitous computing. Related research challenges include power management, software architectures and connectivity with wireless and personal area networks.

Figure 9: Wearable devices, worn by the bearer under, with or on top of clothing but it could be inserted under skin, *source from [30]*.

E. E-Healthcare Systems

Future wireless-equipped healthcare systems are expected to offer high quality services to patients through a variety of applications enabled by information and wireless communication technologies. They can remotely and continuously monitor the patients' health status at home and outdoor. Figure 10 shows the aforementioned technologies; wearable devices and wireless sensors integrated with cellular and wireless systems for patient' services. Early detection of patients' emergency situations makes it possible to provide timely first-aid and access to patients' health information in a pervasive manner, thereby improving both system reliability

and efficiency. The challenges of research and development in the application of wireless communication technologies in healthcare systems include energy-efficiency, Quality-of-Services (QoS), design of healthcare architectures, emergency detection and response, security and privacy, and legal issues.

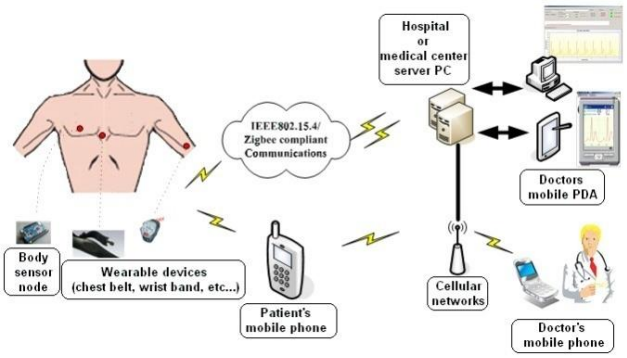


Figure 10: E-healthcare, enabled by wireless and information technologies.

VI. CONCLUSION

System re-configurability based on cognitive radio and software defined radio constitutes one of the main challenges of future research and development for NGNs. Nanotechnology is expected to revolutionize the mobile telecommunications industry where it has significant impacts on both the terminal and core network. All IP network, direct device-to-device communication and Vehicular Ad Hoc Networks are viable for enhancing network' performance and affordability of NGNs. Solid interworking and adaptive coupling of heterogeneous networks are mandatory until one unified global standard for NGNs becomes reality. Global coverage using satellites and high-altitude platforms are possible evolution paths for space roaming and NGN' radio-access extension. Finally, machine-type communication, wireless sensor networks, wearable devices and e-healthcare systems will pave the way for the most developing applications of NGNs.

REFERENCES

[1] Erik Dahlman et al., "4G LTE/LTE-Advanced for Mobile Broadband", Elsevier, 2011

[2] URL: http://www.ipv6.org

[3] S. Haykin, Cognitive radio: Brain-empowered wireless communications, IEEE J. Sel. Area. Comm. 23 (2) (February 2005) 201–220.

[4] Friedrich K. Jondral, "Software-Defined Radio—Basics and Evolution to Cognitive Radio" EURASIP Journal onWireless Communications and Networking 2005:3, 275–283

[5] URL: http://www.sdrforum.org

[6] Leonardo S. Cardoso et al.,"Vandermonde-Subspace Frequency DivisionMultiplexing Receiver Analysis" IEEE 21st International Symposium on, Turkey (2010).

[7] IEEE 802.21: Media Independent Handover,URL: http://www.ieee802.org/21

[8] WWI Ambient Networks.URL: http://www.ambient-networks.org

[9] E. Gustafsson and A. Jonsson, "Always Best Connected," *IEEEWireless Communications Magazine*, February 2003.

[10] IST-FP6 IP MAGNET (My personal Adaptive GlobalNETwork), URL: http://www.telecom.ece.ntua.gr/magnet

[11] IST-FP6 IP MAGNET Beyond (My personal Adaptive Global NET and Beyond), URL: http://www.ist-magnet.org

[12] Liljana M. Gavrilovska, Vladimir M. Atanasovski, "Interoperability in Future Wireless Communications Systems: A Roadmap to 4G", Microwave Review, June, 2007.

[13] TRUST homepage: http://www.ist-trust.org

[14] IST-FP5 MOBIVAS (Mobile Value-Added Services), URL: http://mobivas.cnl.di.uoa.gr

[15] IST-FP5 IP PACWOMAN (Power Aware Communications forWirelessOptiMized personal Area Networks),URL: http://www.imec.be/pacwoman

[16] Qiwei Zhang et al. "A Reconfigurable Radio Architecture for Cognitive Radio in Emergency Networks" Proceedings of the 9th European Conference on Wireless Technology

[17] R.K.Jain, Risal Singh, "Role of Nanotechnology in future wireless and communication systems", National seminar proceeding, Academy of Business & Engineering Science Ghaziabad, pp-19-28, 16-17th January 2009.

[18] Peter Mell and Timothy Grance, "The NIST Definition of Cloud Computing" US National Institute of Standards and Technology Special Publication 800-145, September 2011.

[19] URL: www.cloudcomputingtechnology.org

[20] http://www.vanet.info/

[21] URL: http://www.wimaxforum.org

[22] IEEE 802.16m-07/002r4, TGm System Requirements Document (SRD).

[23] Erik Dahlman et al., "4G LTE/LTE-Advanced for Mobile Broadband", Elsevier, 2011

[24] ITU-R, ITU paves way for next-generation 4 G mobile technologies; ITU-R IMT-advanced 4 G standards to user new era of mobile broadband communications, ITU Press Release, 21 October 2010.

[25] http://www.3gpp.org/release 7

[26] Wireless World Initiative New Radio, Eurescom, 2006, https://www.ist-winner.org.

[27] Psiaki, M. L., "Block Acquisition of weak GPS signals in a software receiver", Proceedings of ION GPS 2001, the 14th International Technical Meeting of the Satellite Division of the Institute of Navigation, Salt Lake City, Utah, September 11-14, 2001, pp. 2838-2850.

[28] High Altitude Platforms for Wireless Communications, T.C.Tozer, D. Grace, Electronics&Communication Engineering Journal June 2001.

[29] http://en.wikipedia.org/wiki/RQ-4_Global_Hawk.

[30] http://www.scienceahead.com/entry/top-21-wearable-technologies/

AUTHORS PROFILE

Currently working as associate professor at the University of Palestine, Dr. Mousa was granted the PhD on "4G Cellular and WLAN Inter-working Networks" with excellent degree from the National Technical University of Athens, Dept. of Electrical & Computer Engineering in 2004. Dr. Mousa obtained a D.E.A (equivalent Master degree) in Digital Telecommunication Systems from {Ecole Nationale Supérieure des Télécommunications/PARIS} in 1996 and the (B.Sc.) of Electronic Engineering from {Middle East Technical University/Ankara} in 1992.

APPENDIX

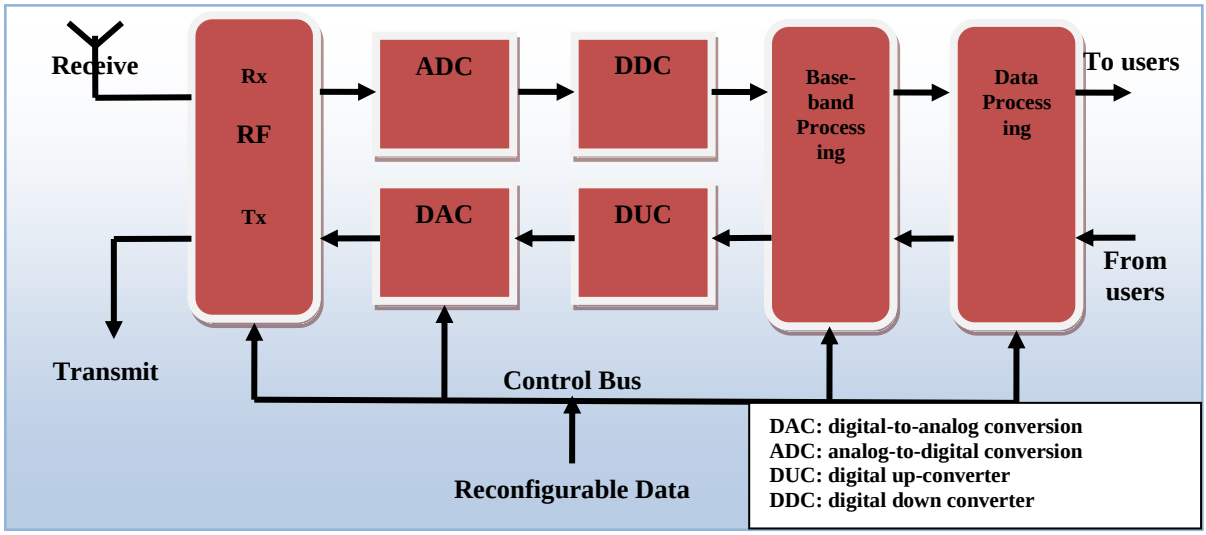


Figure 1: General SDR transceiver

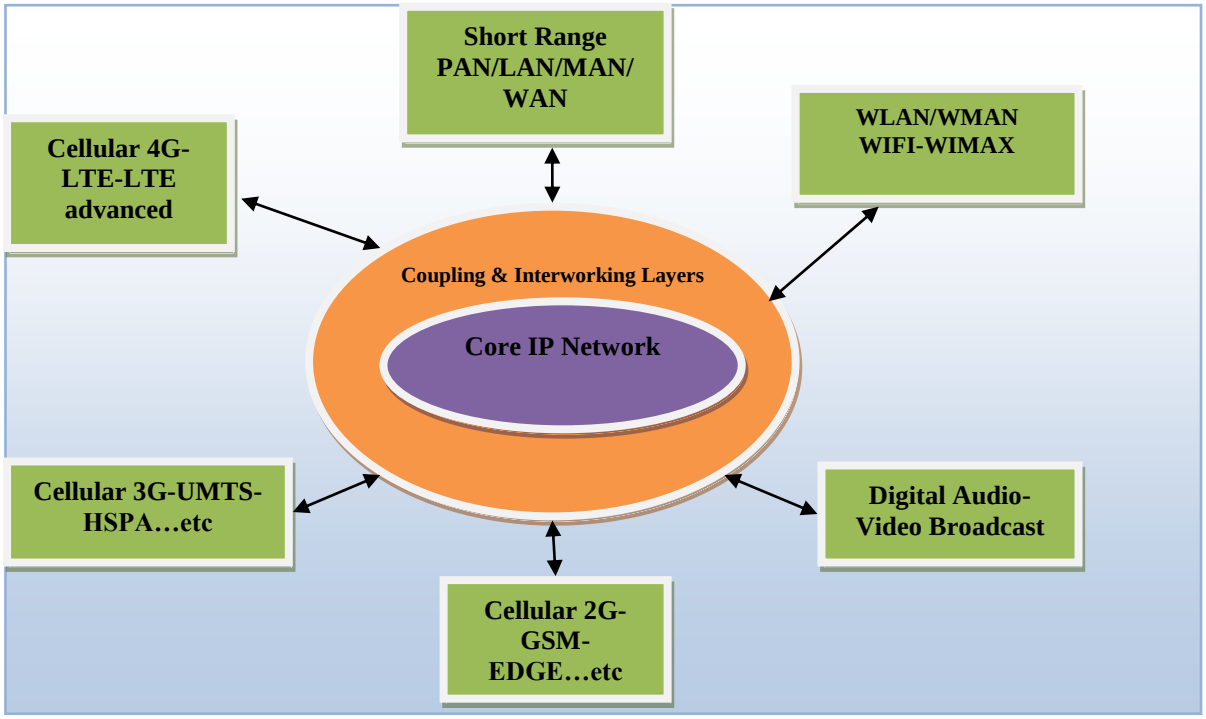


Figure 5: Core IP Network, different access systems' provision.

A Semantics for Concurrent Logic Programming Languages Based on Multiple- Valued Logic

Marion Glazerman Ben-Jacob
Department of Mathematics and Computer Information
Science Mercy College
Dobbs Ferry, New York, USA

***Abstract*— In order to obtain an understanding of parallel logic thought it is necessary to establish a fully abstract model of the denotational semantics of logic programming languages. In this paper, a fixed point semantics for the committed choice, non-deterministic family of parallel programming languages, i.e. the concurrent logic programming languages is developed. The approach is from an order theoretic viewpoint. We rigorously define a semantics for a Guarded Horn Clauses-type of language because of the minimal restrictions of the language. The extension to other concurrent logic programming languages would be direct and analogous, based on their specific rules of suspension. Today's world is replete with multitasking and parallelism in general. The content of this paper reflects a paradigm of an application of multi-valued logic which is reflective of this.**

Keywords- concurrent logic programming; multiple-valued logic; denotational semantics.

I. INTRODUCTION

Parallelism in technology encourages us to examine the meaning of different logical operations being done concurrently. A question of reasonable complexity is how would one attempt to solve more than one quantitative problem at the same time or attempt to make more than one logical inference concurrently. We need to give meaning to the programs that are written in languages that possess the capability of concurrency.

In the case of implementing a logic program via resolution, it appears that not only will parallelism increase efficiency but that the underlying inference procedure actually lends itself in a natural way toward concurrency; for any selected clause, many different instantiations of a selected atom might be attempted at once, more than one atom might be chosen for possible resolution, and any give atom might even select several different clauses that contain an atom against with which it might be resolved. With all these non-deterministic possibilities, obviously, concurrency controls must also be guaranteed by parallel programming languages.

Given that sound results can be obtained, we need to specifically understand how logical inferences are made by machines that support parallelism. A fundamental goal becomes the understanding of parallel logic.

First order logic does not provide us with a significantly sophisticated basis for interpreting parallel logic thought. Perhaps, Kleene's [19] three-valued logic or Belnap's [2] four-valued logic is more appropriate. Upon closer inspection the situation is more involved than just several processors working independently. A reasonable model of what happens when pertinent information is spread over a number of sites that communicate with each other was investigated by Fitting [10, 11]. It was based on Belnap's four-valued logic, exhibited by a bi-lattice structure.

Existing results credit a greater complexity to parallel logic thinking than information just being distributed over a number of sites. Parallelism, accounting for hardware and supporting language, allows for shared memory, sometimes somewhat restricted, sometime global, and interaction among the processors prior to the final assignment of truth values; thus, a fixed point semantics based on a simple bi-lattice structure no longer seems adequate.

In order to obtain an understanding of parallel logic thought and the role played by classical logic, we need to establish a fully abstract model of the denotational semantics of logic programming languages [22]. In addition to its other merits, this type of model can serve as a theoretical foundation for debugging parallel logic programs.

In pursuit of the understanding of parallel logic programming, an analysis of the operational semantics of Concurrent Prolog and the Concurrent Constraint Programming (CCP) family of languages has been made by Saraswat [24]. Kok [20] has developed a purely topological model for the denotational semantics of Concurrent Prolog, and Gerth, Codish, Lichtenstein and Shapiro [14] have developed one for Concurrent Prolog based on sets of suspensions. We will examine the semantics of a parallel logic programming language from an order theoretic viewpoint.

In this paper fixed point semantics for the committed choice, non-deterministic and parallel (CCNAP) family of parallel programming languages, i.e. the concurrent logic programming languages, is developed. We start with GHC, Guarded Horn Clauses, and the simplest of the parallel logic programming languages.

GHC, with its minimal restrictions, few rules of suspension and global environments is an appropriate starting place for understanding how logical implications are formed in parallel. We assume the version of GHC which we are considering allows for failure; thus, it might perhaps be more accurate to say we are considering a GHC-type language. We rigorously define semantics for a parallel programming language. We

build this definition with the operational semantics of GHC in mind and so, this semantics will describe a formalized GHC. The extension to the other concurrent logic programming languages would be natural and based on the language-dependent suspension rules.

II. SYNTAX

The syntax of GHC and more generally, that of any member of the family of CCNAP languages is based on the syntax of the sequential logic programming languages. Let L be a first-order language over a non-empty domain D. A term is either a variable, or a constant or a function of terms, i.e. an element belonging to the domain D. If P( ) is an n-place predicate symbol and t_i , $1 \leq i \leq n$ are terms of L(D) then P($t_1,\ldots,t_n$) is an atom or atomic formula. A literal of L(D) is an atom or the negation of an atom belonging to L(D).

An expression of the form H: $-G_1 \ldots G_k \mid B_1 \ldots B_n$ (k, n>0) is a guarded program clause. H is called the clause head, the G_j's are guard goals and the B_i's are body goals. H, the G_j's and the B_i's are all atomic formulae. The commitment operator, $\mid$, is usually interpreted as conjunction and separates the clause's guard from its body, the former being written to the left of the operator and including the head, and the latter the right. The guard of a non-goal clause is never empty. The limiting case for GHC is when the predicate is the system predicate true, i.e. the clause if of the form h(x): - true $\mid$ b(x). Goal clauses are of the form : - $B_1 \ldots B_n$. A CCNAP program is a finite set of guarded program clauses.

III. SEMANTICS

Before giving formal definitions, we will attempt to provide the motivation for our choice of a five-valued logic and for the actual truth values chosen.

We want our denotational semantics to align with the operational semantics of the CCNAP family of languages as closely as possible, and so, we concluded that we need a four-valued logic, the truth values being true (t), false (f), undefined ($\perp$), and suspend (s). The fifth value, overdefined ($\top$), is included for topological facility.

The necessity of the first two truth values in our logic is obvious. Suspend is required because of the nature of the parallel programming languages. All the concurrent logic languages experience suspension of processing as a result of specific language dependent occurrences. According to Ueda [27], GHC's rules of suspension are:

"(a) Unification invoked directly or indirectly in the guard of a clause C called by a goal G cannot instantiate the goal G.

(b) Unification invoked directly or indirectly in the body of a clause C cannot instantiate the guard of C until that clause is selected for commitment.

A piece of unification that can succeed only by making such bindings is suspended until it can succeed without making such bindings."

The truth value suspend reflects the fact that work has been attempted to establish the truth value of the instantiated predicate in question, no (exact) precise truth valued has yet been assigned and at this point in time work must be stopped and recorded (suspended) so as not to interfere with the validity of the calculations of the other processors. Also, should there be a malfunction in the hardware allowing for one instantiated predicate to be assigned true by one processor and false by another, we will say the predicate has the truth value suspend, indicating that some work has been done on it. We are assuming all processor are working with all clauses.

After careful consideration, it will become clear that $\perp$ belongs in the scheme. We will be determining truth value assignments based on the operations of several processors and our intuition leads to the naturalness of assigning $\perp$ to an instantiated predicate in the following cases: if a processor has not even begun dealing with its values yet; if more "work" is needed before assigning it a truth value and this work can proceed without interfering or contradicting the operation of the other processors.

More formally, along the lines of Fitting and Ben-Jacob [12, 13] we get

Definition 1

FIVE is the space of truth values $\{\top, t, f, suspend, \perp\}$ with the ordering $<_5$ where $\perp <_5$ suspend $<_5$ f $<_5 \top$ and $\perp <_5$ suspend $<_5$ t $<_5 \top$. Figure 1 illustrates the ordering pictorially.

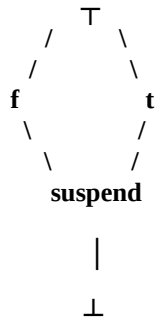


Figure 1

Clearly, the ordering $<_5$ is based on the amount of information or knowledge available. We note the existence of an alternative ordering $<_5^*$ where f $<_5^*$ suspend $<_5^*$ t,

f $<_5^* \top <_5^*$ t and f $<_5^* \perp <_5^*$ t. The ordering $<_5^*$ is based on the amount of truth available.

Definition 2

A five-valued interpretation is a mapping V from ground (variable-free) atomic formulae of L to FIVE. Obviously, in general, an interpretation can be trivially single-valued or many-valued, not necessarily five-valued. For our purposes, we need only consider four-valued interpretations.

Our interpretations are given the point wise ordering based on ground atomic formulae. The space FIVE is a complete lattice and by a generalization of the Knaster-Tarski Theorem, must have a least fixed point and a greatest fixed point [17].

The extension of interpretations from atomic formulae, e.g. A and B, to all closed formulae is governed by the truth tables, Table I and Table II.

TABLE I

A	B	A v B	A∩B
true	false	t	f
	true	t	t
	suspend	t	suspend
	⊥	t	⊥
false	false	f	f
	suspend	suspend	f
	⊥	⊥	f
suspend	suspend	suspend	suspend
	⊥	⊥	⊥
⊥	⊥	⊥	⊥

TABLE II

A	~A
true	false
false	true
suspend	suspend
⊥	⊥

With regard to Table I we note that both AvB and A∩B are monotonic with regard to the knowledge ordering $<_5$ of FIVE. Other patterns that exist include that A∩B takes the g.l.b. of its values with regard to the truth ordering $<_5^*$. A∩B takes the g.l.b. of its values with regard to the knowledge ordering if we disregard the argument being equal to false and lastly, in the knowledge ordering, if one of the arguments is false. A∩B always takes on the l.u.b. of the two values. Also, AvB takes on the l.u.b. of its values with regard to the truth ordering.

Given a clause of the generic type h ← g | b , by definition the guard includes those predicates to the left of the commit operator, including the head. By headless guard we mean only those predicates to the left of the commit operator on the right side of the implication arrow, i.e. the guard without the head of the clause. Consider the following truth table:

TABLE III

headless guard (g)	body b	g | b
t	t suspend f	t suspend f
suspend	t suspend f	suspend suspend
f	t suspend f	f f
⊥	t ⊥	⊥
	f	f
⊥	suspend	suspend

We clarify (truth) Table III by noting if the guard g is false and the body b of a clause suspends, g | b is false since it cannot be used in a proof that ascertains the truth value of a predicate. Also, according to Ueda [27], GHC was designed that with the two given rules of suspension (see page 7), anything can be done in parallel or even executed in a predetermined order provided the latter constraint does not change the meaning of the program; thus, we must allow the possibility of the truth values that appear in the last line of Table III. It is conceivable in theory that the body of a clause could get instantiated prior to the instantiation of the guard and the commitment and that this is the clause to be used in an attempted resolution process. If this takes place, the truth value of the body would be suspend. In this case, if a body, b, suspends its corresponding guard will never be instantiated; thus, the guard of this clause is undefined. Reiterating, if the situation is such that the instantiation of a predicate in the body forces the instantiation of a variable in the guard prior to commitment the truth value of the body would be suspend and so, the guard would never get assigned a truth value. We note that g | b is monotonic. We now define a conditional truth table for the natural interpretation of h ← g | b where g | b is of the form $g_1,\ldots,g_k \mid b_1,\ldots,b_n$ (k,n>0).

Table IV

g | b	h	h← g | b	Annotations
t	t	t	
f	t	t	
suspend	t	t	
⊥	t	t	
t	f	f	
f	f	t	
suspend	f	suspend	Suspended work on g | b caused lack of knowledge of outcome; so, h← g | b has value suspend.
⊥	f	⊥	Insufficient amount of information caused lack of knowledge of outcome, so h← g | b has value
t	suspend	suspend	
f	suspend	suspend	h← g | b takes on suspend to reflect some work was done but not enough to determine a final truth value.
suspend	suspend	suspend	Work has been done on g | b and h, but no value can be determined.
⊥	suspend	suspend	Reflects that work has been done but no conclusion has been reached.
t	⊥	⊥	It is the truth value of h that is responsible for the lack of a determined truth value of the entire clause.
f	⊥	⊥	
⊥	⊥	⊥	

After examining Table IV, we note the monotonicity with regard to knowledge of h←g | b. We also remark on the deviation from classical logic, e.g. a→b ≠ a v b as f → ⊥ is ⊥ and ¬ f v ⊥ = t v ⊥ is true.

Obviously, if the formula is of the form ∃ x A (*there exists an x such that A is true),* its truth value is determined by the truth tables and the relationship ∃ x A = $v_{x\varepsilon D}$ A. Similarly, if the formula is of the form *for- every-x A is true,* its truth value depends on the tables and the relationship *for- every-x A is true* =$\cap_{x\varepsilon D}$ A.

Next we will consider the assignment of truth values for heads of clauses in a logic program, first with regard to an individual processor and then, with respect to several processors working concurrently.

When we write h←g | b we will be using this shorthand notation to mean h(x)←g(x) | b(x). If we want to resolve clauses of the form $h_1(x)$←$g_1(x)$ | $b_1(x)$ and $h_2(y)$←$g_2(y)$ | $h_1(y)$ we would of course first have to unify the clauses. Using our shorthand notation, we would represent these two clauses by h_1←g_1 | b_1 and h_2←g_2 | h_1'.

Let p_1←g_1 | b_1 be the clause whose head predicate p_1 is attempting to unify with the goal clause via a unifier θ_0. Let p_2← g_2 | b_2 be the clause whose head predicate p_2 is attempting to unify with g_1 via a unifier θ_1 to solve g_1. In general, let θ_n be the unifier needed to unify guard g_n with head predicate p_{n+1} to solve the guard g_n. The process is nested j+1 levels until the guard of the j+1st clause is "true."

Definition 3

We will say p and p_i are unifiable if the composition of unifiers $\theta j \cdot \ldots \theta_1 \cdot \theta_0$ does not cause a suspension. The most general unifier will be the composition of most general unifers that do not cause a suspension. Unifiers that do not cause a suspension under composition will be said to be compatible.

Theorem 1

The rigorously defined parallel programming languages that has been defined on the previous pages correctly reflects the operational semantics of GHC.

Proof: Previous discussion, truth tables I-IV, and definitions 2-3.

Definition 4

Let P be a logic program defined over a domain D. A reserved relation symbol is a symbol that represents a given relation on the domain D, i.e. a given mapping from D to the appropriate or relevant space of truth values. Φ_{ip} will be the map on non-parallel interpretations by a given processor i, given by the following: Assume $\Phi_{ip}(V)$ = W. W is the interpretation such that

(i) if R is a reserved relation symbol, W(R(a)) = R(a).

For a non-reserved relation symbol h,

(ii) if ∃ a program clause in P(D) whose head is h(a) and whose guarded body g(a) | b(a) maps to true under V. then W(h(a)) = true.

(iii) if all clauses in P(D) of which h(a) is the head, have guarded bodies that V maps to false, then W(h(a)) = false.

(iv) If at least one clause in P(D) whose head is ha(a) and whose corresponding guarded body V maps to suspend while all other clauses of the form h(a)← $g_i(a)$ | $b_i(a)$ in P(D) have guarded bodies that V maps to false,
W(h(a)) = suspend.

(v) if all clauses in P(D) whose head is h(a) have bodies that V maps to suspend, then W(h(a)) = suspend.

(vi) in all other cases W(h(a)) = ⊥, i.e. the guarded bodies of the clauses in P(D) of which h(a) is the head are such that for at least one, V maps the guarded body to ⊥ and none are mapped to true by V, the W(h(a)) = ⊥.

To gain more comfortable with the previous definition, let us examine its effect on ground programs.

1. Since there does not exist any instantiation of variables, suspension does not play any role as a truth value.
2. Once the headless guard of a clause is true, the interpretation of the clause is equivalent to its "unguarded version." Consider the program

q←
p ←q | r

In this case, V(r) = ⊥, and if Φ(V) = W, then W(p) = ⊥. This program behaves like p←r. Consider the comparison of the following two programs, assuming negation in the body of a clause is allowed.

r←	r←
q←	q←
p←q | r	p←q | ¬ r

In the first program, W(p) = t and in the second program, W(p) = false.

1. If a headless guard is false, V(guarded body) = false; it is irrelevant which truth value V(body) takes on; thus, W(unguarded head) = false.
2. If V(guard) = ⊥, then the truth value of the body is of concern. Consider the following three programs:

h←g | b	b←	b←
	h←g | b	h←g | ¬ b

For the first program V(g | b) = ⊥ and W(h) = ⊥. For the second and third programs, W(h) = ⊥ and W(h) = f, respectively.

The aforementioned conclusions are based on (Truth) Table III. The definition of Φ_p is satisfactory when one processor is trying to interpret the true meaning of a program clause or several processors are concurrently determining the interpretation of one program clause. With parallel logic programming we must account for more than one

interpretation of a clause or a predicate being worked on concurrently, as well.

Definition 5

Let $p_i(a)$ denote the i^{th} interpretation of the instantiated predicate p(a). Then the concept of parallel interpretations is defined in a binary manner and $p_i(a)$ @ $p_j(a)$, the interpretation based on two, concurrent interpretations is given by Table V.

Table V			
$p_i(a)$	$p_i(a)$	$p_i(a)$ @ $p_i(a)$	Annotations
t	t	t	
t	suspend	t	One processor is forced to suspend but the other processor proved the predicate true.
t	f	$\top$	Overdefined- possibly by hardware malfunction; this is the only occurrence of $\top$.
t	$\perp$	t	One processor proved the predicate true; the other one may not even have examined $p_i(a)$.
f	suspend	f	The assumption is that one processor chose the wrong clauses to try to unify and so suspend; the other processor proved the other predicate false based on the program.
f	f	f	
f	$\perp$	f	(Same as other case). One processor did not deal with the truth value of $p_i(a)$ and the other got false as a value.
suspend	suspend	suspend	
suspend	$\perp$	suspend	Some work was done; cannot do more work on one processor and other processor did not work with the predicate.

The commutative and associative closure of @ are obvious. As we previously mentioned the set FIVE with its ordering is a complete lattice (See Figure 1). An interpretation is a map from atomic formulae into the above set. Interpretations with regard to an individual processor are given the point-wise ordering, and any order-preserving map has a least fixed point (lfp) and a greatest fixed point (gfp).

Definition 6

Let Φ denote a parallel-system operator on interpretations based on the maps on the interpretations from the n individual processors with @ defined between operators pointwise.

Lemma 1

Φ is independent of the order of the maps on the individual interpretations upon which Φ is based.

Proof: Referring to Table IV we see the values in the range of the @ operation depends merely on the truth values taken on by the operands under consideration and are order independent.

Interpretations on the system are given a pointwise ordering (i.e. $V_1 < V_2$ iff

$V_1(a) \leq V_2(a)$ with respect to the lattice FIVE for all atoms a) and so, any order-preserving map on parallel interpretations will have a least fixed point. The truth values determined by the lfp of $\Phi_{p,}$, an operator on a five-valued parallel interpretation supplies us with the truth valued determined by the program.

In general, a parallel interpretation is an interpretation that is achieved by parallel evaluation of sequential interpretations. Every sequential interpretation can be considered as (the limiting case of) a parallel interpretation. We now proceed to show the relationship between the parallel interpretations determined by elements in the range of Φ_p and the sequential interpretations determined by the elements in the range of Φ_{ip}, i=1,…n.

Definition 7

For processor i, we define the following family of maps on interpretations with regard to program P. Φ_{ip} is as defined in Definition 4.

(1) $\Phi_{ip}{}^0$ assigns corresponding truth values to reserved relation symbols and given relations; on unreserved relations symbols it is $\perp$.
(2) $\Phi_{ip}{}^{\alpha+1} \equiv \Phi_{ip}(\Phi_{ip}{}^{\alpha})$
(3) For a limit ordinal λ, $\Phi_{ip}{}^{\lambda} \equiv \sup \{ \Phi_{ip}{}^{\alpha} \mid \alpha < \lambda\}$

Definition 8

Let Φ_{ip} be as in Definition 4. Let X be an interpretation of a program P upon which Φ_{ip} is defined. The following are parallel-system interpretations:

$\Phi_p{}^0(X) = X$
$\Phi_p{}^1(X) = \Phi_{1p}{}^1(X) @ \Phi_{2p}{}^1(X) \ldots . @ \Phi_{np}{}^1(X)$
$\Phi_{p_p}{}^{\alpha+1}(X) = \Phi_p(\Phi_p{}^{\alpha}(X))$
$\Phi_p{}^{\lambda}(X) = \text{lub}\ \{\Phi_p{}^{\beta}(X) \mid \beta < \lambda\}$

The following shows $\Phi_p{}^{\alpha}(X)$ is well defined.

Lemma 2

Let μ be any function mapping {1,2,…n} to {0,1} such that lub μ(i) = 1.

Then $\Phi_p{}^1(X) = @_i \Phi_i{}^{\mu(i)}(X)$.

Proof: $\Phi_p{}^1(X)$ is a parallel-system interpretation based on the lub of interpretations implied by sequential interpretations.

Lemma 3:

$\Phi_p{}^{\alpha}(X) = @_i \Phi_i{}^{\mu(i)}(X)$ where μ is any mapping from {1,2…n} to {0,1,2…α} such that $\text{lub}_i(i) = \alpha$, α any ordinal.

Proof: From lemma 1 we see that $\Phi_p{}^{\alpha}(X)$ is a unique interpretation for α a successor ordinal; thus, $\Phi_p{}^{\lambda}(X)$ is well defined for λ a limit ordinal.

Lemma 4:

Let $\Phi_p = @\Phi_{ip}$ be the map from 5-valued parallel interpretations into 5-valued parallel interpretations as defined by Definition 8. Then

(1) Φ_p (@ $\Phi_{ip}^{\lambda i}(X)$) $_{i=1}^{n}$ = (2) (@ $\Phi_{ip}^{\lambda i+1}(X)$) $_{i=1}^{n}$ = (3) @[Φ_p ($\Phi_{ip}^{\lambda i}(X)$)] $_{i=1}^{n}$

Proof: (1) ↔ (2)

(2) is the n-parallel interpretation $\Phi_p^{\alpha}(X)$ where $\alpha = \text{lub}_i (\lambda_i + 1)$. (1) is the interpretation one gets from Φ operating on $\Phi^{\lambda}(X)$, $\lambda = \text{lub}_i \lambda_i$ which by definition = $\Phi^{\lambda+1}(X)$ where

$\lambda = \text{lub}_i \lambda_i$ and so is equal to (2).

(2) ↔ (3)

(3) is the resulting interpretation from Φ_p acting on each sequential interpretation $\Phi_{ip}^{\lambda i}(X)$ as the limiting case of a parallel interpretation, i.e. $\Phi_{ip}^{\lambda i}(X) = \Phi_{ip}^{\lambda i}(X)$ @ $\Phi_{jp}^{0}(X)$ (for all $j \neq i$). When we perform the @ operation we arrive at the parallel interpretation $\Phi_p^{\alpha}(X)$ where $\alpha = \text{lub}_i (\lambda_i + 1)$.

IV. Conclusion

We contend that the approach of defining the semantics of parallel logic programs that we have presented here has a strong relationship with the theory of powerdomains based on the topology established by Plotkin [24]. Additional future work includes an extension of the results of this paper to other concurrent logic programming language and their respective semantics [15].

References

[1] Apt, K.R., Bezem, M., Acrylic Programs, New Generation Computing, 9, pp. 335-363, (1995).

[2] Belnap Jr., N.D., A Useful Four-Valued Logic, Modern Uses of Multiple Valued Logic, (edited by Dunn and Epstein) Reidel, Dordrecht, pp. 8-37, (1977).

[3] Chen, W., Warren, D.S., A Goal-Oriented Approach to Computing the Well-Founded Semantics, Journal of Logic Programming, 17, pp. 279-300, (1993).

[4] Chen, W., Warren, D.S., Toward Effective Evaluation of General Logic Programs, Technical Report 93-CSE-11, Southern Methodist University, (1993).

[5] Della Croce, F., Tsoukiàs A., Moraïtis P., "Why is Difficult to Make Decisions under Multiple Criteria, Proceedings of the Sixth International Conference on AI Planning & Scheduling *(AIPS'02) Workshop on Planning and Scheduling with Multiple Criteria*, pp. 41-45, Toulouse, France, (2002).

[6] Derensart, P. Maluszynski, J. A Grammatical View of Logic Programming, MIT Press, (1993).

[7] Dung, P. an Argumentation Semantics for Logic Programming with Explicit Negation, Proceeding 10th International conference on Logic Programming, pp. 615-30, (1993).

[8] Fitting, M.C., The Family of Stable Models, Journal of Logic Programming, 17, pp. 197-225, (1993).

[9] Fitting, M.C., A Kripke-Kleene Semantics for Logic Programs, Journal of Logic Programming, vol.3 pp. 295-312, (1986).

[10] Fitting, M.C., Logic Programming on a Topological Bilattice, Fundamentica Informatica, vol. 11, pp.209-218, (1988).

[11] Fitting, M.C., Bilattices and the Semantics of Logic Programming, Journal of Logic Programming, vol.11, pp. 91-116, (1991).

[12] Fitting, M.C.,, Ben-Jacob, M., Stratified and Three-Valued Logic Programming Semantics, Logic Programming, Proceedings of the Fifth International Conference and Symposium, editors, Kowalski, R.A., and Bowen, K. S. pp.1054-1069, The MIT Press, (1988).

[13] [13] Fitting, M.C., Ben-Jacob, M., Stratified, Weak Stratified, and Three-valued Semantics, Fundamenta Informatica, vol.13, pp. 19-33, (1990).

[14] Gerth, R., Codish, M., Lichtenstein, Y., Shapiro, E., Fully Abstract Denotational Semantics for Flat Concurrent Prolog, Weizmann Institute Technical Report CS-8803, (1988).

[15] Hewitt, Carl, The repeated demise of logic programming and why it will be reincarnated; What Went Wrong and Why: Lessons from AI Research and Applications. Technical Report SS-06-08. AAAI Press. (March 2006).

[16] Huth, M., Jagadeesan, R., and Schmidt, D.A. Modal Transition Systems: a Foundation for Three-valued Program Analysis. Proceedings of the European Symposium on Programming, Springer LNCS 2028, pp. 155-169, (2001).

[17] Kakas, A., Mancarella, A., Dung, P., The Acceptability Semantics for Logic Programs, Proceedings of the 11th International Conference on Logic Programming, pp.504-519, (1994).

[18] Knaster, B. Une Theoreme sur les Fonctions d'Ensembles, Ann. Soc. Polon. Math. Vol. 6, pp. 133-134 (1928).

[19] Kleene, S.C., Introduction to Metamathematics, Van Nostrand, Princeton, (1952).

[20] Kok, J.N., A compositional Semantics for Concurrent Prolog, Symposium on Theoretical Aspects of Computer Science, pp. 373-388, (1988).

[21] Malfon, B., Characterization of Some Semantics for Logic Programs with Negation and Application to Program Validation, Rapport de Recherche Laboratoire d' Informatique Fondamentale d'Orleans, pp. 94-100, (1994).

[22] Milner, Robin, The Space and Motion of Communicating Agents. Cambridge University Press, (2009).

[23] Ross, K.A., A Procedural Semantics for Well-Founded Negation in Logic Programs, Journal of Logic Programming, 13, pp. 1-22, (1992).

[24] Saraswat, V.A., Concurrent Constrain Programming Languages, Ph.D. thesis, Carnegie-Mellon University, (January 1989).

[25] [25] Schmidt, David, Denotational Semantics, Wm. Brown, Iowa, (1988).

[26] Tarski, A., A Lattice-Theoretical Fixpoint Theorem and its Applications, Pacific Journal of Mathematics, vol.5, pp. 285-309, (1955).

[27] Ueda, K. Guarded Horn Clauses, ICOT Technical Report TR-103, June 1985.

[28] Ueda, K. and Kato, N., The Language Model LMNtal. Proceedings of the 19th International Conference on Logic Programming (ICLP'03), LNCS 2916, Springer-Verlag, pp.517-518, 2003.

[29] Ueda, K. A Pure Meta-Interpreter for Flat GHC, A Concurrent Constraint Language, *Computational Logic: Logic Programming and Beyond* (Essays in Honour of Robert A. Kowalski, Part I), A.C. Kakas, F. Sadri (Eds.), Lecture Notes in Artificial Intelligence 2407, Springer-Verlag, pp.138-161, (2002).

[30] Van Gelder, A. Ross, K.A., Schlipf, J.S., The Well Founded Semantics for General Logic Programs, Journal of the ACM, vol. 38, 3, pp. 620-650, (1991).

Authors Profile

Dr. Marion Ben-Jacob is a Professor in the Department of Mathematics and Computer Information Science at Mercy College for over 30 years. She teaches computer science, mathematics, and critical inquiry, both in the traditional classroom and online. She has recently written papers and spoken extensively on the topics of computer science, computer ethics, online teaching/distance education, collaborative learning, and global learning at numerous conferences. Dr. Ben-Jacob serves on the editorial board of the Journal of Educational Technology Systems. She is the editor and a contributing author of Integrating Computer Ethics across the Curriculum and a recently published e-book, Computer Ethics: Integrating across the Curriculum.

A Decision Tree Classification Model for University Admission System

Abdul Fattah Mashat
Faculty of Computing and Information Technology
King Abdulaziz University
Jeddah, Saudi Arabia

Mohammed M. Fouad
Faculty of Informatics and Computer Science
The British University in Egypt (BUE)
Cairo, Egypt

Philip S. Yu
University of Illinois, Chicago, IL, USA
King Abdulaziz University
Jeddah, Saudi Arabia

Tarek F. Gharib
Faculty of Computing and Information Technology
King Abdulaziz University
Jeddah, Saudi Arabia

***Abstract*— Data mining is the science and techniques used to analyze data to discover and extract previously unknown patterns. It is also considered a main part of the process of knowledge discovery in databases (KDD). In this paper, we introduce a supervised learning technique of building a decision tree for King Abdulaziz University (KAU) admission system. The main objective is to build an efficient classification model with high recall under moderate precision to improve the efficiency and effectiveness of the admission process. We used ID3 algorithm for decision tree construction and the final model is evaluated using the common evaluation methods. This model provides an analytical view of the university admission system.**

Keywords- Data Mining; Supervised Learning; Decision Tree; University Admission System; Model Evaluation.

I. INTRODUCTION

Data mining, the science and technology of exploring data in order to discover unknown patterns, is an essential part of the overall process of knowledge discovery in databases (KDD). In today's computer-driven world, these databases contain massive quantities of information. The accessibility and abundance of this information make data mining a matter of considerable importance and necessity [1].

Data mining includes many methods and techniques, but mainly we can divide them into two main types; verification and discovery. In verification-oriented methods, the system verify the user's input hypothesis like goodness of fit, hypothesis testing and ANOVA test. On the other hand, discovery-oriented methods automatically find new rules and identify patterns in the data. Discovery-oriented methods include clustering, classification and regression techniques.

Supervised learning methods attempt to discover the relationship between input attributes and target attribute. Once the model is constructed, it can be used for predicting the value of the target attribute for a new input data. There are two main supervised models: classification models, which is our interest in this paper, and regression models. Classification models build a classifier that maps the input space (features) into one of the predefined classes. For example, classifiers can be used to classify objects in an outdoor scene image as person, vehicle, tree, or building. While, regression models map the input space into real-vales domain. For example, a regression model can be built to predict house price based on its characteristics like size, no. of rooms, garden size and so on.

In data mining, a decision tree (it may be also called Classification Tree) is a predictive model that can be used to represent the classification model. Classification trees are useful as an exploratory technique and are commonly used in many fields such as finance, marketing, medicine and engineering [2, 3, 4, 5]. The use of decision trees is very popular in data mining due to its simplicity and transparency. Decision trees are usually represented graphically as a hierarchical structure that makes them easier to be interpreted than other techniques. This structure mainly contains a starting node (called root) and group of branches (conditions) that lead to other nodes until we reach leaf node that contain final decision of this route. The decision tree is a self-explanatory model because its representation is very simple. Each internal node test an attribute while each branch corresponds to attribute value (or range of values). Finally each lead assigns a classification.

Fig. 1 shows an example for a simple decision tree for "Play Tennis" classification. It simply decides whether to play tennis or not (i.e. classes are Yes or No) based on three weather attributes which are outlook, wind and humidity [6].

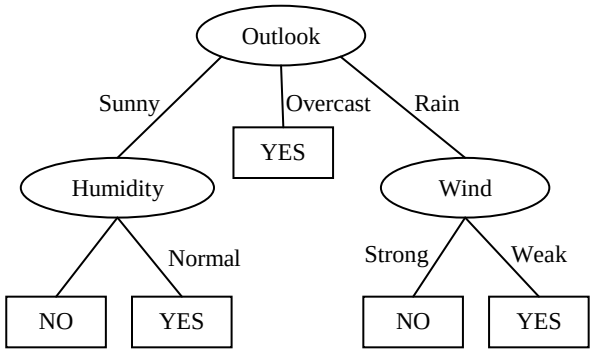


Figure 1. Decision Tree Example.

As shown in Fig. 1, if we have a new pattern with attributes outlook is "Rain" and wind is "Strong", we shall decide not to play tennis because the route starting from the root node will end up with a decision leaf with "NO" class.

In this paper, we introduce a supervised learning technique of building a decision tree model for King Abdulaziz University (KAU) admission system to provide a filtering tool to improve the efficiency and effectiveness of the admissionprocess.KAU admission system contains a database of records that represent applicant student information and his/her status of being rejected or accepted to be enrolled in the university. Analysis of these records is required to define the relationship between applicant's data and the final enrollment status.

This paper is organized into five sections. In section 2, the decision tree model is presented. Section 3 provides brief details about commonly used methods for classification model evaluation. In section 4, experimental results are presented and analyzed with respect to model results and admission system perspective. Finally, the conclusions of this work are presented in Section 5.

II. Decision Tree Model

A decision tree is a classifier expressed as a recursive partition of the input space based on the values of the attributes. As stated earlier, each internal node splits the instance space into two or more sub-spaces according to certain function of the input attribute values. Each leaf is assigned to one class that represents the most appropriate or frequent target value.

Instances are classified by traversing the tree from the root node down to a leaf according to the outcome of the test nodes along this path. Each path can be transformed then into a rule by joining the tests along this path. For example, one of the paths in Fig. 1 can be transformed into the rule: "If Outlook is Sunny and Humidity is Normal then we can play tennis". The resulting rules are used to explain or understand the system well.

There are many algorithms proposed for learning decision tree from a given data set, but we will use ID3 algorithm due to its simplicity for implementation. In this section we will discuss ID3 algorithm for decision tree construction and some of the frequently used functions used for splitting the input space.

A. ID3 Algorithm

ID3 is a simple decision tree learning algorithm developed by Quinlan [7]. It simply uses top-down, greedy search over the set of input attributes to be tested at every tree node. The attribute that has the best split, according to the splitting criteria function discussed later, is used to create the current node. This process is repeated at every node until one of the following conditions is met:

- Every attribute is included along this path.
- Current training examples in this node have the same target value.

Fig.2 shows the pseudo code for ID3 algorithm to construct a decision tree over a training set (S), input feature set (F), target feature (c) and some split criterion (SC).

B. Splitting Criterion

ID3 algorithm uses some splitting criterion function to select the best attribute to split with. In order to define this criterion, we need first to define entropy index that measures the degree of impurity of the certain labeled dataset.

For a given labeled dataset S with some examples that have n (target values) classes {c1, c2, …, cn), we define entropy index (E) as in (1).

$$E(S) = \sum_{i=1}^{n} p_i * \log(p_i), \quad p_i = \frac{|S_{c_i}|}{|S|} \qquad (1)$$

Where S_{c_i} the subset of the examples that have a target value that equals to c_i. Entropy (E) has its maximum value if all the classes have equal probability.

$ID3(S, F, c, SC)$
Output: Decision Tree T

Create a new tree T with a single root node
IF no more split (S) **THEN**
Mark T as a leaf with the most common value of c a label.
ELSE
$\forall f_i \in F$ find f that has best $SC(f_i, S)$
Label t with f
FOR each value v_j of f
Set $Subtree_j = ID3(S_{f=v_j}, F - \{f\}, c, SC)$
Connect node t to $Subtree_j$ with edge labeled v_j
END
END
Return T

Figure 2. ID3 Algorithm

1) Information Gain

To select the best attribute for splitting of certain node, we can use information gain measure, Gain (S, A) of an attribute A, bya set of examples S. Information gain is defined as in (2).

$$Gain(S, A) = E(S) - \sum_{v \in V(A)} \frac{|S_{A=v}|}{|S|} E(S_{A=v}) \qquad (2)$$

Where E(S) is the entropy index for dataset S, V(A) is the set of all values for attribute A.

2) Gain Ratio

Another measure can be used as a splitting criterion which is gain ratio. It is simply the ratio between information gain value Gain(S, A) and another value which is split information SInfo(S, A) that is defined as in (3).

$$SInfo(S, A) = \sum_{v \in V(A)} \frac{|S_{A=v}|}{|S|} * \log \frac{|S_{A=v}|}{|S|} \qquad (3)$$

3) Relief Algorithm

Kira and Rendell proposed the original Relief algorithm to estimate the quality of attributes according to how well their values distinguish between examples that are near to each

other [8]. The algorithm steps are stated in Fig. 3, where *diff* function calculates the difference between the same attribute value (A) within two different instances I1 and I2 as in (4).

$$diff\ (A, I_1, I_2) = \begin{cases} 0 & I_1[A] = I_2[A] \\ 1 & \text{otherwise} \end{cases} \quad (4)$$

Relief

Input: Training set S with N examples and K attributes

Output: Weights vector W for all attributes A

Set all weights $W[1..K] = 0$

FOR i = 1 to N

Select random example **R**.

Find nearest hit **H** (instance of the same class).

Find nearest miss **M** (instance of different class).

FOR A = 1 to K

$$W[A] = W[A] - \frac{diff\ (A,R,H)}{N} + \frac{diff\ (A,R,M)}{N}$$

END

END

Return W

Figure 3. Relief Algorithm

III. Model Evaluation

Consider a binary class problem (i.e. has only two classes: positive and negative), the output data of a classification model are the counts of correct and incorrect instances with respect to their previously known class. These counts are plotted in the confusion matrix as shown in table 1.

TABLE I. Confusion Matrix (Binary Class Problem)

True Class	Predicated Class		
	Positive	**Negative**	
Positive	TP	FN	CN
Negative	FP	TN	CP
	RN	**RP**	**N**

As shown in table 1, TP (True Positives) is the number of instances that correctly predicted as positive class. FP (False Positives) represents instances predicted as positive while their true class is negative. The same applies for TN (True Negatives) and FN (False Negatives). The row totals, CN and CP, represent the number of true negative and positive instances and the column totals, RN and RP, are the number of predicted negative and positive instances respectively. Finally, N is the total number of instances in the dataset.

There are many evaluation measures used to evaluate the performance of the classifier based on its confusion matrix resulted from testing. We will describe in more details some of the commonly used measures to be used later in our experiment.

Classification Accuracy (***Acc***) is the most used measure that evaluates the effectiveness of a classifier by its percentage of correctly predicted instances as in (5).

$$Acc = \frac{TP + TN}{N} \quad (5)$$

Recall (***R***) and Precision (***P***) are measures that are based on confusion matrix data. Recall (R) is the portion of instances that have true positive class and are predicted as positive. On the other hand, Precision (P) is the probability of that a positive prediction is correct as shown in (6).

$$R = \frac{TP}{CN} \text{ and } P = \frac{TP}{RN} \quad (6)$$

Precision and recall can be combined together to formulate another measure called "***F-measure***" as shown in (7). A constant β is used to control the trade-off between the recall and the precision values. The most commonly used value for β is 1 that represents F1 measure.

$$F_\beta = \frac{(1+\beta^2) * P * R}{(\beta^2 * P) + R} \quad (7)$$

For all the defined measures above, their values range from 0 to 1. For a good classifier, the value of each measure should reach 1.

Another common evaluation measure for binary classification problems is ***ROC*** curve that is firstly proposed by Bradley in [9]. It is simply a graph that plots the relation between the false positive rate (x-axis) and true positive rate (y-axis) for different possible cut-points of a diagnostic test. The curve is interpreted as follows:

- The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test.
- The closed the curve comes to the 45o diagonal of the ROC space, the less accurate the test.
- The area under the ROC curve measures overall accuracy. An area of 1 represents a perfect test, while an area of 0.5 represents a worthless test.

IV. Experiments

A. Dataset

King Abdulaziz University (KAU) admission system in the Kingdom of Saudi Arabia (KSA) is a complex decision process that goes beyond simply matching test scores and admission requirements because of many reasons. First, the university has many branches in KSA for both division male and female students. Second, the number of applicants in each year is a huge which needs a complex selection criterion that depends on high school grades and applicant region/city.

In this paper, we are provided by sample datasets from KAU system database that represent applicant student information and his/her status of being rejected or accepted to be enrolled in the university in three consecutive years (2010, 2011 and 2012). The dataset contains about 80262 records, while each record represents an instance with 4 attributes and the class attribute with two values: Rejected and Accepted. The classes are distributed as 53% of the total records for "Rejected" and 47% for "Accepted" class. Table 2 shows detailed information about datasets attributes.

The dataset is divided into two main parts: training dataset that holds about 51206 records (about 64%) and testing dataset that contains about 29056 records (about 36%). The decision tree classifier is learnt using a training dataset and its performance is measured on not-seen-before testing datasets.

TABLE II. SUMMARY OF DATASET ATTRIBUTES

Attribute	Possible values
Gender	Student's gender • Male • Female
HS_Type	Type of high school study • TS = Scientific Study • TL = Literature Study • TU = Unknown/Missing
HS_Grade	High school grade • A = mark ≥ 85 • B = 75 ≥ mark > 85 • C = 65 ≥ mark > 75 • D = 50 ≥ mark > 65
Area	Code for student's region city (116 distinct value)

B. Decision Tree Model Results

The decision tree model is generated over training dataset records using Orange data mining tool [10]. The generated decision tree is a binary tree with "One value against others" option. The confusion matrix valuesare shown in table 3. The values of confusion matrix are generated by applying a decision tree on testing datasets.

TABLE III. TESTING CONFUSION MATRIX

True Class	Predicated Class		
	Accepted	**Rejected**	
Accepted	12305	1538	**13843**
Rejected	8484	6729	**15213**
	20789	**8267**	**29056**

TABLE IV. MODEL EVALUATION MEASURES

Measure	Value
Accuracy	$Acc = \frac{12305 + 6729}{29056} = 0.655$
Recall	$R_{\text{Accepted}} = \frac{12305}{13843} = 0.889$ $R_{\text{Rejected}} = \frac{6729}{15213} = 0.442$
Precision	$P_{\text{Accepted}} = \frac{12305}{20789} = 0.592$ $P_{\text{Rejected}} = \frac{6729}{8267} = 0.834$
F_1 Measure	$F1_{\text{Accepted}} = \frac{2 * 0.592 * 0.889}{0.592 + 0.889} = 0.711$ $F1_{\text{Rejected}} = \frac{2 * 0.834 * 0.442}{0.834 + 0.442} = 0.578$

The evaluation measures shown in table 4 shows that the proposed classifier achieved a high recall at the cost of moderate precision. This means that a filtering tool improved the efficiency and effectiveness of the admission process. The classifier is to filter out the low level candidates so the admission staffs can focus their energy on the most promising candidates to make a better selection. So, the workload on the administrative staff is much reduced and hence they may be able to make a better selection job. In fact missing some (i.e., With a recall slightly lower than 1) is not necessarily bad, as the administrative staffs may not always be able to identify the best candidates from a large pool. On the other hand, the same measures in case of "Rejected" class are about 0.58. This mid-level value stated that the classifier performance is above average.

C. Decision Tree Induced Rules1

One of the main advantages of the decision tree is that it can be interpreted as a set of rules. These rules are generated by traversing the tree starting from the root node till we reach some decision at a leaf. These rules also give a clear analytical view of the system under investigation. In our case, they will help KAU admission system office to understand the overall process. The induced set of rules is stated in table 5.

TABLE V. DECISION TREE RULES SET

IF Area = "1007" **AND** HS_Grade = "A" **THEN** "Accepted" (75.7%)
IF Area ≠ "1007" **AND** HS_Grade = "A" **AND** Gender = "Male" **AND** Area = "1001" **THEN** "Accepted" (74.9%)
IF Area ≠ "1007" **AND** HS_Grade = "A" **AND** Gender = "Female" **AND** Area ≠ "901" **THEN** "Rejected" (64.4%)
IF Area ≠ "1007" **AND** HS_Grade = "A" **AND** Gender = "Female" **AND** Area = "901" **THEN** "Rejected" (85.0%)
IF Area ≠ "1007" **AND** HS_Grade ≠ "A" **AND** HS_Grade ≠ "B" **THEN** "Rejected" (98.9%)
IF Area ≠ "1007" **AND** HS_Grade = "A" **AND** Gender = "Male" **AND** Area ≠ "1001" **THEN** "Rejected" (51.1%)
IF Area ≠ "1007" **AND** HS_Grade ≠ "A" **AND** HS_Grade = "B" **THEN** "Rejected" (90.5%)
IF Area = "1007" **AND** HS_Grade ≠ "A" **AND** HS_Grade ≠ "B" **THEN** "Rejected" (87.0%)
IF Area = "1007" **AND** HS_Grade ≠ "A" **AND** HS_Grade = "B" **THEN** "Rejected" (63.9%)

As shown in table 5, beside each rule there is the percentage of instances that have the predicted class by this rule. Also, we can figure out that there are only two rules that lead to "Accepted" state. The first occurs if the student area code is "1007" (which is "Jeddah" city) and student's high school grade is "A" (which is excellent student). The second case when "Male" student from area with code "1001" (which is "Rabigh" city) with grade "A" in high school.

V. CONCLUSION

In this paper we presented an efficient classification model using decision tree for KAU university admission office. The experimental results show that a filtering tool improved the efficiency and effectiveness of the admission process. This is

achieved by the decision tree classifier with high recall under moderate precision (which determines the candidate pool size).we induced a set of rules by using the decision tree structure that helps KAU admission officeto make a better selection in the future.

The model stated that the most accepted students from "Jeddah" region in KSA with excellent high school grade (more than 85%) or male students from "Rabigh".

REFERENCES

[1] J. Han and M. Kamber, (2000)," Data mining:concepts and techniques", San Francisco, Morgan-Kaufma.

[2] H.S. OH and W.S. SEO, (2012), "Development of a Decision Tree Analysis model that predicts recovery from acute brain injury", Japan Journal of Nursing Science. doi: 10.1111/j.1742-7924.2012.00215.x

[3] G. Zhou and L. Wang, (2012), "Co-location decision tree for enhancing decision-making of pavement maintenance and rehabilitation", Transportation Research: Part C, 21(1), 287-305. doi:10.1016/j.trc.2011.10.007

[4] S. Sohn and J. Kim, (2012). "Decision tree-based technology credit scoring for start-up firms: Korean case ", Expert Systems With Applications, vol. 39(4), 4007-4012. doi:10.1016/j.eswa.2011.09.075

[5] J. Choand P.U. Kurup, (2011), "Decision tree approach for classification and dimensionality reduction of electronic nose data", Sensors & Actuators B: Chemical, vol. 160(1), 542-548.

[6] T. Mitchel, (1997), Machine Learning, USA, McGraw Hill.

[7] J. R. Quinlan, (1986), "Introduction of Decision Tree", Machine Learning, vol. 1, pp. 86-106.

[8] K. Kira and L.A. Rendell (1992), "A practical approach to feature selection", In D.Sleeman and P.Edwards, editors, Proceedings of International Conference on Machine Learning, pp. 249-256, Morgan Kaufmann.

[9] A.P. Bradley, (1997), "The use of the area under the roc curve in the evaluation of machine learning algorithms", Pattern Recognition, vol. 30, pp. 1145-1159.

[10] Orange Data Mining Tool: http://orange.biolab.si/

A semantic cache for enhancing Web services communities activities: Health care case Study

Hela Limam
Department of Computer Science
High Institute of Management
Tunis, Tunisia

Jalel Akaichi
Department of Computer Science
High Institute of Management
Tunis, Tunisia

***Abstract*— Collective memories are strong support for enhancing the activities of capitalization, management and dissemination inside a Web services community. To take advantages of collective memory, we propose an approach for indexing a health care Web services community's' resources with semantic annotations explaining and formalizing its informative content. Then we show how the health care Web services community' members exploit their collective memory by expressing their queries allowing them searching relevant resources in order to perform their activities.**

Keywords- Web services community; semantic description; health care.

I. INTRODUCTION

Despite their visible advantage and accessibility, the rapid growing number of published Web services prevents users or requestors from finding easily and efficiently the services relevant to their specific needs. Hence, the concept of communities of Web services has emerged for gathering Web services according to their functionalities in order to ease and improve the process of Web services discovery in an open environment like the Internet. By providing a centralized access to several functionally-equivalent Web services via a unique endpoint, communities enable processing complex users' queries that a single Web service cannot satisfy.

The cornerstone of building Web services communities is their ability to be queried transparently and easily by users, which aim to satisfy their informational needs in a satisfactory time and in a pertinent retrieval. Nevertheless, processing a user query is not an easy task and may involve the access to a number of distributed communities in order to locate Web services that are capable of answering the query. Those queries are sometimes complex and short-living. Hence, it seems to be beneficial to conserve them for a future reuse and in order to be shared by communities' members who have similar informational needs.

In this context, collective memories appear to be very attractive to use in order to enhance sharing useful dedicated reusable fragments of know-how inside a Web services community. Enhancing Web services communities activities using a semantic cache memory highlights the interest of capitalizing formulated queries to the cache memory and in general the expert how-know of the community in the field of the information discovery. Hence, the process of researching resources for answering queries becomes based on a formal manipulation of annotated resources.

In our paper we propose a model for Health care services communities which enables semantic caching of queries, we provide a formal description of queries and the cache content and we detail query processing inside a community using the semantic cache.

The rest of this paper is organized as follows: In Section 2, we review previous research on semantic caching as well as other related issues. A Health care community model is defined in Section 3. Section 4 proposes a semantic cache model suitable for Web services communities. Section 5 discusses the semantic cache organization and the semantic caching query processing strategies. Finally, we summarize our work and discuss future research in Section 6.

II. RELATED WORKS

A review of research projects aiming at assisting community activities by a collective memory as the european project SevenPro [1], the projects ANR e-WOK HUB [2] and Immunosearch [3] or the projet C3R [4] highlights the need to capitalize requests made by users in suitable databases to allow their authors to reuse or exchange them with other community members. More generally, capitalization approaches to information retrieval becomes a real issue in many areas.

Indeed, specific strategies are implemented by experts in a specific field in order retrieve information necessary for their activities [5] and are often difficult to acquire by novice users. These strategies, more and more critical with the increasing specialization and knowledge bases are capitalized and are rarely used in either the search tools like Google or in the portal domain [5].authors in [5] propose an approach to clarify the critical procedures of information retrieval in the medical field using what they call the search strategies portals .Starting with a set of standard questions in the field, they define a set of patterns representing research procedures. A search procedure is represented by an ordered set of subgoals and for each search procedure, links to relevant sources of information are established

Authors [6] offer a browsing environment of Web resources. They distinguish among three levels of knowledge: (i) a medium level of knowledge that brings web resources in the field of application (ii) Represents a level of knowledge

that brings together the meta-data resources of the previous level and (iii) a level of knowledge transmission that offers courses (called "e-course") for web resources through the meta-data associated with them. E-courses are composed of steps characterized by an intention or a title, subject and an illustration. The illustrations are Web resources related to e-course step.

Examples of research procedures for dynamic navigation systems also exist in online learning-based models of semantic Web. In [7], a model for the pedagogical approach is adopted by an assembly of requests parameterized and resources whose annotations meet these requests up educational material are presented dynamically to the learner navigating through the system.

In conclusion we can say that there has been much interest in the area of applying semantic caching in Web services and communities in general. Some of the proposed approaches only work in the field Web communities while others limit queries to Web services. Hence previous works either ignore the possibility of applying semantic cache for enhancing and performing query processing among communities of Web services .The objective of our work is to extend the existing semantic caching work along several dimensions. First we present a formal semantic caching model for Web services communities, and then we explore the semantic caching query processing strategies. We examine how to efficiently answer queries against the cache. Finally, we validate semantic cache performance through a detailed Health care case study.

III. HELTH CARE WEB SERVICES COMMUNITIES MODEL

Communities of Web services are virtual spaces that can dynamically gather different Web services having complementary functionalities in order to provide composite services with high quality. Some approaches have been proposed to organize communities of Web services. In a previous work we proposed a Web services communities design language, called WSC-UML[12], which increases the expressiveness of UML for Web services communities and guides their design. Stereotypes and graphical annotations have been added to UML diagrams in order to distinguish between the different aspects in a Web services community. WSC-UML was used to model Web services communities in general. The studied formalism is suitable for modeling a Health care community in a way that enables querying Health care Web services.

In this section we introduce a WSC-UML model for a Health care community and we specify its associated Web services. The Health care community is designed to combine data from the large ancillary services, such as pharmacy, laboratory, and radiology, with various clinical care Services .It has an identifier and is described by a set of attributes. It is composed of a set of Health care Web services: Insurance Service, Care Service, Patient Refferal Service, Physician Refferal Service and scheduling Service. Web services insider a community are associated to each other's with peer relationships. Each Web service modeled as a class in WSC-UML class diagram and is described by a set of attributes as shown in figure 1. The number of integrated Web services involved in the Health care community is dependent upon the data structures and has to provide an interface that allows clinicians to access the silo systems through a portal.

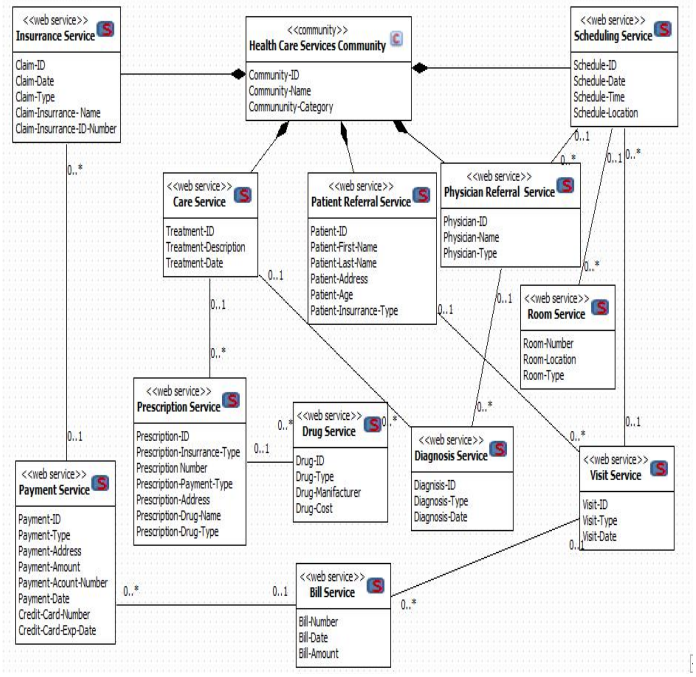


Figure 1. WSC-UML Class Diagram of a Health care community

IV. A SEMANTIC CACHE MODEL FOR WEB SERVICES COMMUNITIES

A. The system general architecture

We propose a model for building, reusing and sharing queries for a pertinent information retrieval. The proposed model takes as input queries expressed by users then transforms them into a format that allows their reuse. We focus on the construction the community semantic cache that can be seen as episodic memories in which the research approaches are built dynamically based on queries. We also focus on the scenario of the information retrieval that we call the process of finding information. The general architecture of our system is exposed in figure 2.

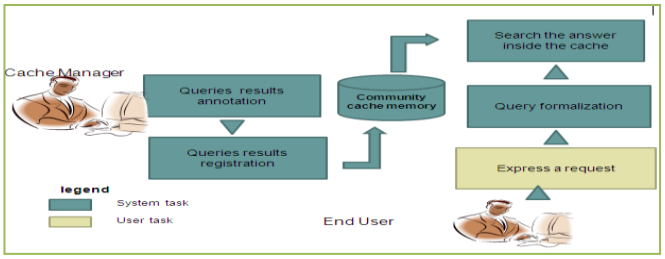


Figure 2. The semantic cache general architecture

B. Formal definition of the semantic cache content

The queries formulation in our work is partially inspired by the work [8]. Our approach distinguishes itself by the fact that it extends and adapts the cited work in the field of Web service communities. We consider that queries addressed to communities take the form of select and project queries, although the proposed model can be extended to handle other

more complicated queries .the formal definition of the semantic cache is given in this section, later we discuss how to store and organize the semantic cache. The Web services community model presented in the previous section is considered a relational database where each modeled class is a relation. For example the class Web service is a relation. Each relation consists of a relation schema and a relation instance. An instance of the relation << Community Member>> is a set of members satisfying the constraint of having the same number of the attributes described in the Community Member schema.

Suppose that the examined community C consists of a set of modeled classes cl1, cl2,.., cln

$$C=\{ cli , 1\leq i \leq n \}$$

Further let Acli stand for the attribute set defined by the schema of the class cli and A present the attribute set of the whole community, then we have:

$$A=\cup A_{cli},\ 1\leq i \leq n$$

Before defining a semantic cache, we first give the predicates, through which a semantic cache is constructed.

Definition 1: Given a community $C=\{cl_i\}$ and its attributes set $A=\cup A_{cli},\ 1\leq i \leq n$, *a* **Compare Predicate** *of* C, P, is of the form P=a op c,

where a ϵ A, op ϵ {≤, <, ≥,>,=},c is domain value or a constant.

In fact, a semantic cache is used to store annotated results of queries. It is composed of a set of Semantic Query Result. A Semantic Query Result is an original, decomposed, or coalesced query result. Its definition is consistent with that of a materialized view [9]. To further simplify the problem, we assume that the selection condition of a query is an arbitrary constraint formula of compare predicates, namely, a disjunction of conjunctions of compare predicates.

Definition 2: Given a community $C=\{cl_i\}$ and its attributes set $A=\cup A_{cli}, 1\leq i\leq n$, a Semantic Query Result is a tuple $<C_C , S_A , S_P, S_C>$ where $S_C=\pi_{SA}\sigma_{SP}(C_C)$, $S_R \epsilon C$, $S_A \subseteq A_{CC}$ and $S_{P=} P_1 \vee P_2 \vee \ldots P_m$ where each P_j is a conjunctive of comparative predicates, i.e., $P_j=b_{j1}\wedge b_{j2}\wedge b_{jl,}$

Each b_{jl} is a compare predicate involving only the attributes in $A_{CC.}$

In definition 2, CC and SA define the class and attributes involved in computing S, respectively, SP indicates the select condition that the tuples in S satisfy. Hence, these three elements specify the semantic information associated with S. The actual content of S is represented by SS. From the restrictions added, we can see that semantic query results are the results of Select-Project operations, with the selection conditions containing only compare predicates. Before queries get answered, their contents are empty (i.e., QC=Φ). Therefore, we formally define a query just as we define a semantic query result.

Definition 3: A Query Q has the form $Q= <Q_C,Q_A,Q_P ,Q_S >$ A semantic cache is defined as a set of semantic queries results. To reduce space overhead, the cached semantic queries results do not overlap with each other. In the following, we first give the concept of disjointed queries results, and then formally define a semantic cache.

Definition 4: Two Semantic queries results $S_i =<S_{iC} ,S_{iA} ,S_{iP} ,S_{iS} >$ and $Sj =<S_{jC} ,S_{jA} ,S_{jP} ,S_{jS} >$ are said to be disjointed if and only if :

1. $S_{iA}\cap S_{jA} =\Phi$ or

2. $S_{iP} \wedge S_{jP}$ is unsatisfiable.

Definition 5: A Semantic Cache, S C,is defined as SC={semantic query result S_i} where ∀ j, k(Sj ϵSC ∧ Sk ϵ SC ∧ j≠ k) Sj and Sk are disjointed).

C. Semantic cache organization

Several approaches tackle the problem of the physical storage of semantic queries results. The work in [10]stores queries results in tuples, and associates every query with a pointer to a linked list of the corresponding tuples. This approach works fine for select-only queries and memory caching. The key advantage is easy maintenance: tuples can be added, deleted, or moved between segments conveniently. However, this linked list scheme is not appropriate for disk caching, since it may result in too many I/O operations.

Moreover, when select-project queries are cached, the resulting tuples for different segments are no longer at the same length. Hence, even for memory caching, its advantage in maintenance is lost. Another noticeable disadvantage for this approach is the large space overhead caused by the tuple pointers.

In our case, semantic cache is composed of two parts: the content and the index. Every semantic query result is stored in one or multiple linked pages, and is associated with a pointer pointing to its first page in the memory (disk) cache. Each page contains a query result, rather than the community classes. The cache space is also managed at a page level, which makes semantic query result allocation and deallocation algorithms more straightforward and simpler.

For allocation, if there are enough free pages to hold a query result, and then allocate the pages to it; for deallocation, just mark the deallocated pages as free. The index part maintains the semantic as well as physical storage information for every cached query result. In what follows, we list the basic items kept in the index. For every cached query result, we have:

- The name S, the community class C_C, the attribute set S_A, and the selection predicate S_P
- The pointer pointing to the first page that stores the query result
- The timestamp indicating when the query result was last visited S_{TS}

The index structure proposed here is consistent with the formal definition of the semantic cache. In addition to the four basic components of the semantic query result, we further add other items for maintenance use, such as S_{TS}. The semantic cache index is more clearly illustrated through the following Example 1.

Example 1: Consider a health care community with two classes: Patient referral service and scheduling services

Patient referral service (Patient-ID, Paddress, Ptelephone, Pfirst-name, Plast-name, P Age, Pinsurance-Type) and

Scheduling Services (Schedule-ID,Sdate, Stime,Slocation); also suppose that the cache contains four queries result:

- S_1: Select Plast-name From Patient referral service Where 20 < PAge < 60;
- S_2: Select Pfirst-name, Plast-name From Patient referral service Where PAge> 10;
- S_3: Select Schedule-ID From Scheduling Services Where SDate= 28/12/2010;
- S_4: Select Slocation From Scheduling Services Where Sdate > 10/12/2010 ;

Also, suppose that the first pages of the four queries results are one, three, five, and six, respectively, and S_i was last visited at T_i, then the index is shown in Table 1.

TABLE 1. QUERIES RESULTS ORGANIZATION IN SEMANTIC CACHE

S	C_C	S_A	S_P	S_C	S_{TS}
S1	Patient referral service	Plast-name Pfirst-name,	20 < PAge < 60 PAge> 10	1	T_1
S2	Patient referral service	Plast-name Schedule-ID	SDate= 28/12/2010	3	T_2
S3	Scheduling Service	Slocation	Sdate > 10/12/2010	5	T_3
S4	Scheduling Service			6	T_4

V. SEMANTIC CACHING AND QUERY PROCESSING

To process a query from a semantic cache, we first check whether it can be answered by the cache. If yes, the locally available results are computed directly from the cache. When the query can only be partially answered, we trim the original query by removing or annotating the already answered parts and send it to the database server for processing. In this section, algorithms for semantic caching query processing are examined.

A. Theoretic Foundation

From the concept of Derivability defined in [11] we introduce the following definition.

Definition 6: Consider a semantic query result $S=<C_C,S_A,S_P,S_C>$ and a query $Q=<Q_C,Q_A,Q_P,Q_S>$,we say Q is answerable from S, if there exist a relational algebra expression F containing only project and select operations, and only involving attributes in S_A, such that $F(S_C)\neq\Phi$ and $\forall t\ (t\epsilon F(S_C)) \Rightarrow$ (t satisfies $Q_P \wedge$ t contains only attributes in Q_A)). Furthermore, if $F(S_C)=Q_C$, we say Q is fully answered from S; otherwise, we say Q is partially answered from S.

From Definition 6, we know that the key to compute a query from a cached segment is to find the function F, and to make sure that F can be executed on the segment. Sometimes, even the entire result of a query Q is contained in a segment S, Q still is not answerable from S.

This is because some of the attributes needed in F cannot be found in S. So, in Definition 6, we add an additional restriction on F. The following Example 2 illustrates such a point.

Example 2: Consider health care community and the semantic cache described in Example 1, suppose there comes a query Q = Select Pfirst-name, Plast-name From Patient referral service Where(PAge> 5) ∧ (Patient-insurrance-type= Personal). Obviously, the result of Q is totally contained in S_2, since every tuple which satisfies (PAge > 10) ∧ (Patient-insurrance-type= Personal) will always satisfy (PAge > 5). However, Q cannot be computed from S_2, because we cannot find an F as specified in Definition 6.

Intuitively, Q seems to be computed from S_2 by a function π Pfirst-nameσ (PAge > 10) ∧ (Patient-insurrance-type= Personal)

But the attribute"Patient-insurrance-type" used in this function is not in S_2 after the projection.

Definition 7: Consider a query $Q=<Q_C,Q_A,Q_P,Q_S>$, the predicate attribute set, Q_{PA} , contains all the attributes that occur in Q_P , i.e., Q_{PA} = { a |a is an attribute, and a occurs in Q_P}. Consider a semantic query result $S=<C_C,S_A,S_P,S_C>$, a query $Q=<Q_C,Q_A,Q_P,Q_C>$, and Q's predicate attribute set Q_{PA} . Then we have:

- Statement 1: If $C_C = Q_C$, $S_A \cap Q_A \neq\Phi$, $Q_P \wedge S_P$ is satisfiable by C_C, and $Q_{PA} \subseteq S_A$, then Q is answerable from S.
- Statement 2: If $C_C = Q_C$, $Q_A \subseteq S_A$, $Q_P \Rightarrow S_P$, and and $Q_{PA} \subseteq S_A$, then, Q is fully answered from S.

Definition 8: Consider a semantic segment $S=<C_C,S_A,S_P,S_C>$. Let Y be the set of all attributes uniquely determined by the attributes in the attribute set X, with respect to SP.If $X \subseteq SA$, we say S is an extensible semantic segment, $S_A \cup Y$,denoted by S_A^+ is called the extended attribute set of S, and the semantic query result $S=<C_C,S_A,S_P,S_C>$ is called the extended query result of S. Since S_A is uniquely determined by X with respect to S_P , if a tuple consisting of attributes in X satisfies S_P , when extended to contain attributes in S_A^+, it will also satisfy S_P .

This makes it possible to extend S to S^+ Notice that for each extensible query result, there could exist multiple different extended attribute sets, and hence multiple different extended squeries results. To investigate how to use extended segments in query processing, we examine Example 2 again. Suppose "Pfirst-name" is the key of "Patient referral service" class, thus other attributes of "Patient referral service" can be uniquely determined by "Pfirst-name" Hence, S_2 is an extensible query result. Clearly, "Patient-insurrance-type" can be uniquely determined by "Pfirst-name" To form S_2^+, we retrieve the tuples containing both "Pfirst-name" and "Patient-insurrance-type" from the community knowledge base append

them to S_{2C} according to the value of "Pfirst-name". After that, Q can be computed from S_2^+.Therefore, we have:

Statement 3: Consider an extensible query result $<C_C,S_A,S_P ,S_C >$ and a query $<Q_C,Q_A,Q_P ,Q_S >$ and suppose SA+ is an extended attribute set of S, and S+ is the extended query result of S with respect to SA+, QPA is Q's predicate attribute set. Then, we have:

1. If $C_C= Q_C$, $S_A^+ \cap Q_A \pm\Phi$, $Q_P \wedge S_P$ is satisfiable,and $Q_{PA} \subseteq S_A^+$, then Q is answerable by S^+.

2. If $C_C= Q_C$, $Q_A \subseteq S_A$, $Q_P \Rightarrow S_P$, and $Q_{PA} \subseteq S_A^+$, then Q can be fully answered by S^+.

Definition 9: Consider a semantic segment $S=<C_C,S_A,S_P ,S_C >$, suppose K_A is the primary key of C_C. If $K_A \subseteq SA$, we say S is a key-contained query result.

Definition 10: Consider a key-contained query result

$S=<C_C,S_A,S_P ,S_C >$, and a query $=<Q_C,Q_A,Q_P ,Q_S >$ and suppose Q_{PA} is its predicate attribute set. Then, we have:

1. If $C_{C=} Q_C$, $Q_P \wedge S_P$ is satisfiable, then Q is answerable by $S+< C_C, S_A,Q_A \cup Q_{PA},S_P ,S^+_C >$.

2. If $C_{C=} Q_C$, Q_P ,S_P , then Q can be fully answered by $S^+ = < C_C, S_A,Q_A,Q_{P A},S_P ,S^+_C >$.

B. Query processing

Since a community does not store Web services locally, processing the query requires locating Web services that are capable of answering the query. These Web services can be selected from the local members of the community or from the semantic cache. We propose a collaborative query processing technique that consists of two steps:

- Dividing the query into parts when put together, satisfy all constraints expressed in the query
- Resolving the query by sending it to the selected parts.

For the first step, we adopte a query rewriting algorithm, which takes as input the community classes $S=<C_C,S_A,S_P ,S_C >$, and the query $Q =<Q_C,Q_A,Q_P ,Q_S >$ then produces the following output:

- Qlocal: the part of the query Q that can be answered by the community's local semantic queries results S, that is, the attributes specified in the query that are supported by the local members. It also gives the combination of the local members that can answer all (or part of) the query.
- Qrest: the part of the query that cannot be answered by the local queries results. The community will identify any external members who can answer this part of the query. Hence, Qrest is forwarded to peers. The expected answers of the forwarding is the combination of the external members that are capable of answering Qrest.

Relationship between Q and S fall into five types as described in figure 3.

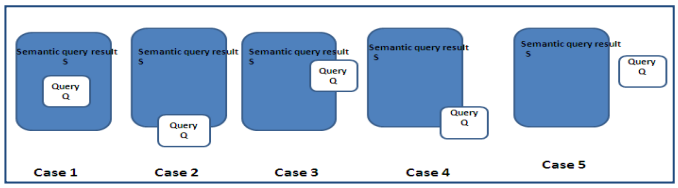


Figure 3. Query and semantic query result relationships

To summarize the query rewritwing work discussed in this section, we present Query Rewriting Algorithm, which rewrites a query Q via a key-contained semantic query result S which is defined on the same relation as Q.

QUERY REWRITING ALGORITHM

Query Rewrit (Query Q, Seamtic query result S, Query lq, Query aq, Query rq1, Query rq2, int type), to rewrite a query Q via a semantic query result S.

Input: Query Q; key-contained semantic query result S

Output: local Query lq; Amending Query aq; Rest Query rq1, rq2; Tr Type type

Procedure: {

KA←S's key attribute set;

A1←(QA∩SA) $\cup K_A$;A2←(QA-SA)∪KA;

IF QA⊆SA {

IF QP⇒SP {

/***** Case 1*****/

type = 1;

IF (QPA⊆SA) THEN aq = NULL; ELSE

aq = πKA σQP (QR)□;

lq = πQA∪KA σQP (SC)□

rq1 = rq2 = NULL; return;}

IF (QP∧ SP is satisfiable) {

/***** Case 2*****/

type = 2;

IF (QPA⊆SA) THEN aq = NULL; ELSE

aq = πKAσ QP∧SP (QR)□;

lq = πQA∪KA σ QP (SC)□;

rq1 = πQA∪KA σ QP∧¬SP (QR);

rq2 = NULL; return;} }

IF (QA⊆SA) does not hold {

IF (QP⇒SP) {

/***** Case 3 *****/

type = 3; lq =π A1(SC)□; rq1 =π A2 σQP (QR)□;

rq2 = aq = NULL; return; }

IF (QP∧SP is satisfiable) {

/***** Case 4 *****/

type = 4;

lq = πA1(SC);rq1=πQAπKA σ QP∧¬SP(QR)

rq2 = πA2σQP∧SP (QR)□; aq = NULL; return; } }

/***** Case 5 *****/

rq1 = Q; \ pq = aq = rq2 = NULL; type = 5; return;}

VI. CONCLUSION

We have presented a semantic cache mechanism designed for enhancing querying a Health care community. Semantic caching is based on the semantic representation of cached data and processing queries by construction of local queries for retrieving cached data and rest queries for fetching data from remote servers. Hence, we proposed semantic cache architecture for caching multiple queries addressed to the community and considered all operational cases. For all types of answers we have developed algorithms for query evaluation against the cache content. In next works we tackle the problem of Web service synchronization when changes occur on Web services which may alter queries results stored in the semantic cache. We also plan to propose replacement strategies for the cache maintenance in order to tune them better to real user profiles.

REFERENCES

[1] Cherif H,Corby O, Faron C, Khelif K," Semantic annotation of texts with RDF graph contexts" In Proceeding of the International Conference on Conceptual Structures (ICCS'2008), pp.75 -82, 2008.

[2] Khalid Belhajjame, Mathieu d'Aquin, Peter Haase, Paolo Missier, "Semantic hubs for geographical projects". In Proceeding of Semantic Metadata Management and Applications (SeMMA), workshop at ESWC, pp. 3–17, 2008.

[3] Kefi L.,Demarkez M,Collard M, "A knowledge base approach for genomics data analysis" In Proceeding of the International Conference on Semantic Systems, Graz, Austria, 2008.

[4] Faron C,Mirbell I,Sall B, Zarli, "Une approche ontologique pour formaliser la connaissance experte dans le modèle du contrôle de conformité en construction " ,19ème journées francophones d'ingénierie des connaissances, Nancy, France,Capaudes Editions ,2008.

[5] BhavnaniS, Bichakjian C, JhonsonT, Little R, Peck F, Strecher V, "Strategy hubs: Next generation domain protals with search procedures". In Proceeding of ACM Conference on Human Factors in

[6] Buffereau B, Duchet P, Picouet P, "Generating guided tours to facilitate learning from a set of indexed resources". In Proceeding of IEEE International Conference on Advanced Learning Technologies (ICALT), pp. 492, Athens, Greece: IEEE Computer Society, 2003.

[7] Yessad A, Faron C, Dieng R, LASKRI M, "Ontology-driven adaptive course generation for web-based education". In World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED MEDIA), Vienna, Austria, 2008.

[8] Limam Hela, Akaichi Jalel, Oueslati Wided "WSC-UML: A UML Profile for Modeling Web Services Communities: A Health Care Case Study" International Journal of Advanced Research in Computer Science, Vol.2, No.2, Mars 2011.

[9] Gupta , I. Singh Mumick, "Maintenance of Materialized Views: Problems, Techniques, and Applications," Data Eng. Bull.,vol. 18, no. 2, pp. 3-18, 1995.

[10] S. Dar, M.J. Franklin, B.T. Jonsson, D. Srivatava, M. Tan, "Semantic Data Caching and Replacement," In Proceeding of VLDB Conf.,pp. 330-341, 1996.

[11] P.A. Larson, H.Z. Yang, "Computing Queries from Derived Relations" In Proceeding of Very Large Databases, pp. 259-269, 1985.

[12] Limam, H., Akaichi, J., Oueslati, W:WSC-UML: A UML Profile for Modeling Web services communities. Vol. 2, No,2, pp; 285-290,(2011).

MDSA: Modified Distributed Storage Algorithm for Wireless Sensor Networks

Mohamed LabibBorham
Department of Computer Science
Faculty of Computers and Information
Helwan University

Mostafa-Sami Mostafa
Department of Computer Science
Faculty of Computers and Information
Helwan University

Hossam Eldeen Moustafa Shamardan
Department of Information TechnologyFaculty of Computers and Information
Helwan University

***Abstract*—In this paper, we propose a Modified distributed storage algorithm for wireless sensor networks (MDSA). Wireless Sensor Networks, as it is well known, suffer of power limitation, small memory capacity,and limited processing capabilities. Therefore, every node may disappear temporarily or permanently from the network due to many different reasons such as battery failure or physical damage. Since every node collects significant data about its region, it is important to find a methodology to recover these data in case of failure of the source node. Distributed storage algorithms provide reliable access to data through the redundancy spread over individual unreliable nodes. The proposed algorithm uses flooding to spread data over the network and unicasting to provide controlled data redundancy through the network. We evaluate the performance of the proposed algorithm through implementation and simulation. We show the results and the performance evaluation of the proposed algorithm.**

Keywords- Distrubited storage; encoding; decoding; flooding; multicasting; unicasting.

I. INTRODUCTION

Along with the industrial development and technological progress, monitoring the environment plays a very vital role. Many research works have been built-up on monitoring systems that can replace traditional systems in critical environments.

Wireless sensor networks (WSNs) are deployed to an area of interest to sense phenomena. Wireless sensor network is a type of ad-hoc networks that has the ability of sensing and processing data collected from the environment [1]. These networks are comprised of autonomous devices, called sensor nodes. Each sensor has a buffer which can be divided into small slots.Wireless Sensor Networks are recently applied in many environmental applications such as measuring temperature, humidity, salts or monitoring objects and others. WSNs applications have common task, which is environmental monitoring. This task is realized by using nodes to sense data from the environment and sends it to the base station. In all applications, we must take into account to use an efficient data collection approach.

Wireless sensor networks have many advantages but the main problem is the limitation of node's resources. Due to that fact, any node may fail to communicate with other nodes in the network or disappear from the network due to accidental events in harsh environments or due to battery depletion. In many applications, every sensor has important data thatareused to form the total overview about the environment so it is important to find an efficient method to recover the lost sensed data.

The Distributed Storage Algorithm (DSA-I) [2] was introduced as a model for recovering data from the failed nodes in WSNs. DSA-I algorithm depends on using flooding and multicasting to disseminate data packets over all network nodes. Flooding with multicasting cause a crowded messages through the network that rise power overhead, growth memory usage, increase-processing load on the nodes beside the increaseof both packet loss rate, and latency.

In this paper, we propose an unusual methodology which is called a Modified Distributed Storage Algorithm (MDSA) to overcome the previous algorithm DSA-I problems . MDSA algorithm depends on unicasting instead of multicasting to provide controlled-redundancy of data through the network. A consequence of these modifications, the number of created messages and the energy consumption are both reduced (See Table.1). Also, taking into account the buffer status whether it is empty or not. If it is full, it will cancel the operation to reduce processing load on the nodes. The proposed model is applied in large-scale wireless sensor network where these nodes are randomly distributed in the serviced environment.

The rest of this paper is organized as follows. In section 2, we present a short overview of distributed storage techniques and networking codes. In section 3, we introduce the algorithm design, consideration and assumptions of the proposed model. In section 4, we introduce our proposed MDSA algorithm. In section 5, we show a simulation of MDSA and comparison between the proposed algorithm and the initial algorithm DSA-I. In section 6, we explain our simulation results and performance evaluation. Finally, we conclude and present future work.

II. RELATED WORK

Different types of networking codes allow every node in a network to achieve some computation. Therefore, each node is not used only to store data, but also has some processing capabilities to do some function or mixture when the data reach its destination; by that, the encoding process occurs inside the network and finally decoded at the final destination [3].

Using network codes provides reliable data access through redundancy over the nodes [4]. In [5], two schemes for maintaining redundancy-using erasure coding Maximum-Distance Separable (MDS) and Regenerating Codes (RC). In [6], it is used a decentralized implementation of Fountain codes that uses geographic routing and every node has to know its location. The motivation for using Fountain codes instead of using random linear codes is that Fountain codes need O(k ln k) decoding complexity but random linear codes and Reed-Solomon codes (RS) use O (k^3) decoding complexity where k is the number of data blocks to be encoded.

In [7], a technique is presented to increase data persistence in wireless sensor networks, which is called "Growth codes". This technique increases the amount of information that can be recovered at the sink node. "Growth codes" is a linear technique in which information is encoded in an online-distributed way with increasing degree. The authors defined persistence of a sensor network as "the fraction of data generated within the network that eventually reaches the sink node ".They showed that Growth codes can increase the amount of information that can be recovered at any storage node at any time period, whenever there is a failure in some other nodes. The motivation for their work is the positions of the nodes are not identified. In addition, a sensor node cannot know the position of other nodes. They assume around time of updating the nodes, meaning with increasing the time, the degree of a symbol is increasing. This is the idea behind growth degrees. They provide practical implementation of Growth codes and compare its performance with other codes.

In [7], the authors studied the question "how to retrieve historical data that the sensors have gathered even if some sensors are destroyed or disappeared from the network?" They analyzed techniques to increase "persistence" of sensed data in a randomly distributed wireless sensor network. They proposed two decentralized algorithms using Fountain codes to guarantee the persistence and reliability of cached data on unreliable sensors. They used random walks to disseminate data from a sensor (source) node to a set of other storage nodes. The first algorithm Exact Decentralized Fountain codes (EDFC) introduces lower overhead than naive random walk, while the second algorithm Approximate Decentralized Fountain Codes (ADFC) has a lower level of fault tolerance than the original centralized Fountain code, but consumes much lower dissemination cost.

In [7], the authors also proposed a novel decentralized implementation of Fountain codes in wireless sensor networks in an efficient and scalable fashion. The authors did not use routing tables to dissimilate data from one sensor to a set of sensors. The reason was the sensors did not have enough energy or memory to maintain routing table, which is scalable to the size of the network.

In [2 and 8], the authors presented a model for distributed storage algorithms for wireless sensor networks where K sensor nodes (sources) want to disseminate their data to N storage nodes with less computational complexity. They proposed two distributed storage algorithms that used flooding, Lt (Luby Transform) codes.Ltcodesare a special class of Fountain codes to guarantee the persistence of data and random walks to disseminate data between sensors. They also assumed in the first algorithm DSA-I that the total number or sources and storage nodes are known. The second algorithm DSA-II assumed that the total number or sources and storage nodes are not known. In other words, every node in the network can know only the number of neighboring nodes; also can estimate the number of sources and the total number of nodes.

III. DESIGN CONSIDERATION AND ASSUMPTIONS

In this section, we show the model assumptions, we assume that all the nodes in the network are identical in all capabilities and act as sensing and storage node. All nodes are distributed randomly and uniformly in an environment, no node maintains routing or geographic tables. Every node can send a flooding message to the neighboring nodes; Also every node can discover the total number of neighbors by broadcasting a query message, and whoever replies to this message will be a neighbor of this node. All nodes have storage buffer. The buffer size is assumed 10 % of the network size. Every node prepares a packet with its ID, sensed data, hop counter, and a flag that is set to zero or one.

IV. MDSA ALGORITHM

We will present a Modified Distributed Storage Algorithm (MDSA) for wireless sensor networks, where all nodes are able to sense and store data. The algorithm consists of two main operations:

A. Data separation and storage.

B. Data collection

A. Data separation and storage

Data separation and storage operations consist of three main phases: Network initialization, packet preparing and flooding phase, finally Data storage and Unicast. In the coming sections we will describe each phase in sequence.

1) Network Initialization phase:

Wireless sensor network applications are used in many environments that may be dynamic environment. In a dynamic environment, the location of the nodes is not fixed and for any environmental reason the nodes may be moved. In the proposed algorithm assumptions, we assume that at the beginning of running, the network topology is not known and the nodes do not maintain routing or geographic tables. Therefore, each node floods query message in its range, and every node that reacts is considered as a neighbor node. At the end of this phase, every node can count the total number of the neighboring nodes.

2) Packet preparing and flooding phase:

After the initialization phase, each node senses data about the required range, and then preparing one packet that is stored in the first slot in its buffer. This packet consists of four fields:

Packet= (ID, Sensed data, hop_count, Flag)

ID shows the sensed node ID, every node has a unique ID, and it is ordered according to its order of creation. The second field presents the sensed data from the environment and it is different in size according to the wireless sensor application

purpose. The third field presents the hop count, which shows how many times the packet, will be traveled or transferred in the network; every node calculates its hop count according to the number of neighboring nodes. The hop_count is calculated according to the following equation:

hop_count=⌊ The total number of network nodes / the total number of neighboring nodes ⌋.

The last field is a flag to show if the packet is new or old, zero for old data and one for updated packet. After preparing packets, each node floods its packet to all neighboring nodes. See Fig.1.

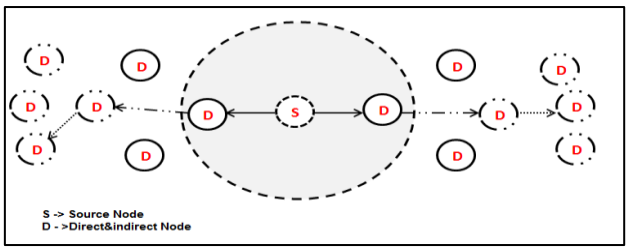


Figure.1 A WSN is randomly distributed in a field. A node **S** is source node and flood its data to its neighboring nodes D then D unicast its data to one of its neighbors and so on.

3) Data storage and Unicast phase:

After flooding of the packets from all the source nodes, the neighboring nodes store the received packet in its buffer if it has a location, and check the hop count if it is greater than zero, the received node will be unicasted the packet to one node from its neighboring nodes. This is repeated until the hop count equal zero the packet will be discarded. Fig.2. Show the MDSA Algorithm Flowchart where S is source node and Iis intermediate node.

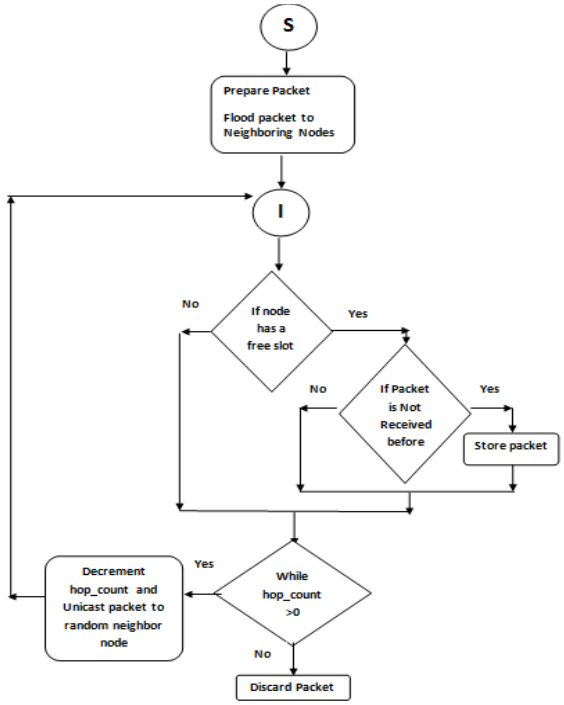


Figure.2 Modified Distributed Storage Algorithm Flowchart (MDSA)

B. Data collection operations

Data separation and storage operations provide data redundancy in the network so the stored data can be recovered by querying a number of living nodes from the network.

V. MDSA SIMULATION AND COMPARISON

We simulate the proposed algorithm MDSA using well-known wireless sensor network simulator called OMNET++ 4. The simulation results show that the performance of the proposed algorithm is better than the previous algorithm DSA-I [2]. The comparisonmetricsare considering many factors, such as energy consumption, number of created messages, memory usage (see Fig.3), data dissemination and Recovering.

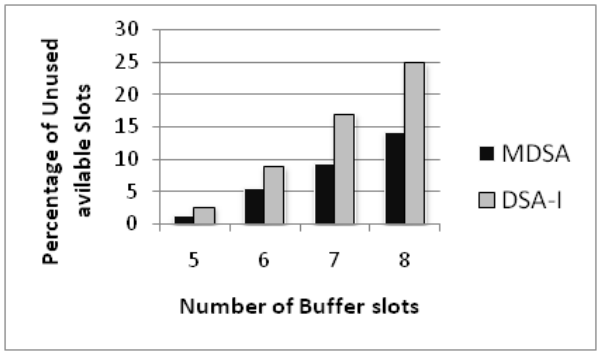


Figure.3 percentage of unused available buffer slots in MDSA and DSA-I algorithm

In Table.1, it is shown the practical comparison between the proposed algorithm MDSA and DSA-I [2] regarding to the number of created messages and percentage of unused buffer's slots with fixed number of nodes = 15, where M is the number of slots in every node's buffer.

TABLE I. COMPARISON BETWEEN MDSA AND DSA-I ALGORITHM

Algorithm	Total Number of Nodes-*N*	No of Created Messages	Percentage of unused buffer's slots			
			M=5	*M*=6	*M*=7	*M*=8
DSA-I (old)	15	29257	2.7%	8.9%	17.1%	25%
MDSA (New)	15	76	1.3%	5.6%	9.5%	14.2%

VI. PERFORMANCE EVALUATION AND SIMULATION RESULTS

After simulating the Modified distributed storage algorithm using OMNET++ 4 simulator, the main performance metric to be investigated, is the average of the successful decoding percentage versus the decoding ratio.

We define the Average of successful decoding percentage as a percentage of recovered data to the total data in the network. In addition, we can define the decoding ratio as the number of nodes, which are queried, divided by the total number of network nodes.

Fig.4 to Fig.9 shows the effect of changing the total number of the network nodes. Moreover, we ran the experiments for

50,100,150,200,400,600 nodes with fixed ratio to the buffer size to be 10% of the network size.

We ran the experiment for many times reach to 30 times to evaluate the performance with various decoding ratios depending on the total number of nodes inside the network with incremental step = 0.1 , at every step we calculate the mean and the standard deviation to make sure that our results are closer to the actual implementation of the algorithm.

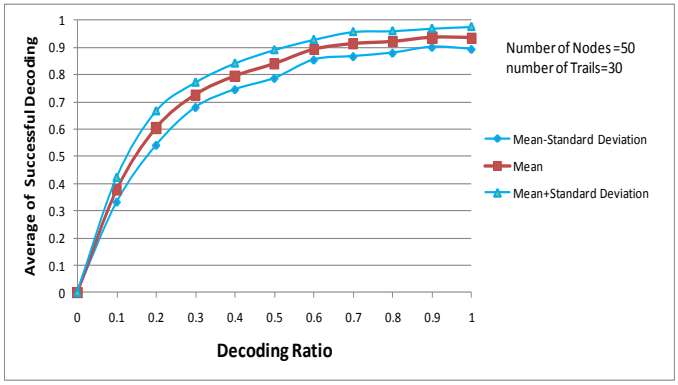


Figure.4.A WSN is randomly distributed in a field. The average of successful decoding ratio is shown for network size=50 nodes and buffer size= 5 slots with the MDSA algorithm

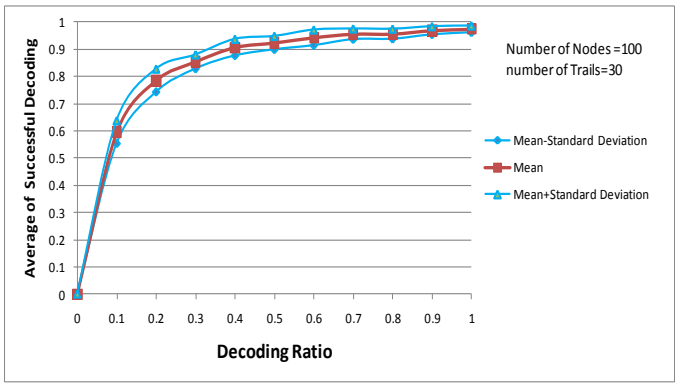


Figure.5.A WSN is randomly distributed in a field. The average of successful decoding ratio is shown for network size=100 nodes and buffer size= 10 slots with the MDSA algorithm

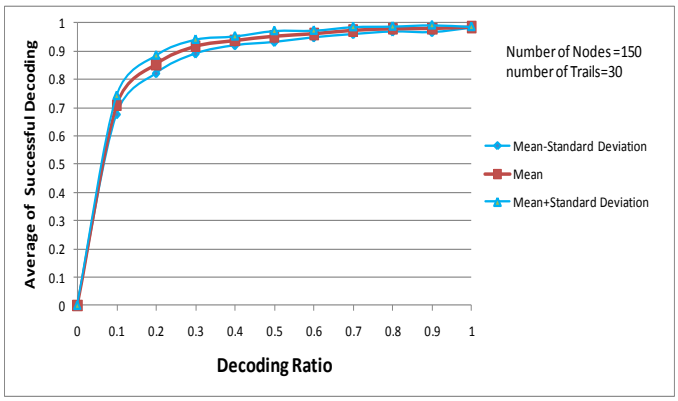


Figure.6.A WSN is randomly distributed in a field. The average of successful decoding ratio is shown for network size=150 nodes and buffer size= 15 slots with the MDSA algorithm

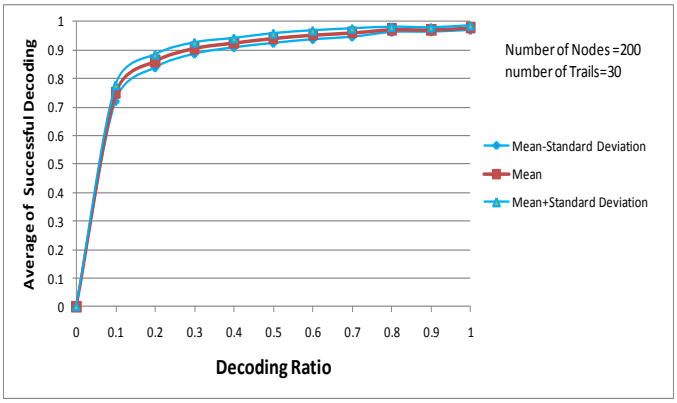


Figure.7.A WSN is randomly distributed in a field. The average of successful decoding ratio is shown for network size=200 nodes and buffer size= 20 slots with the MDSA algorithm

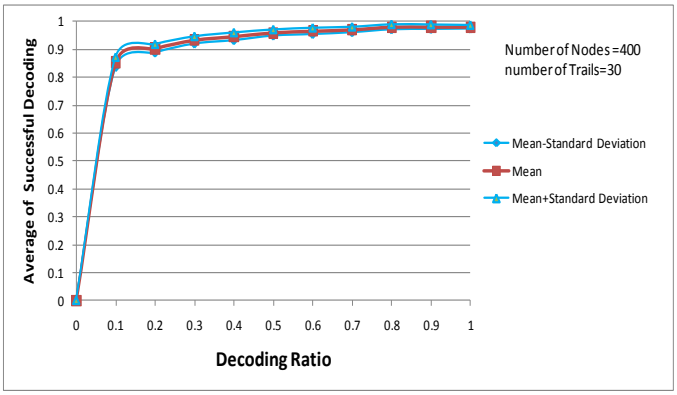


Figure.8.A WSN is randomly distributed in a field. The average of successful decoding ratio is shown for network size=400 nodes and buffer size= 40 slots with the MDSA algorithm

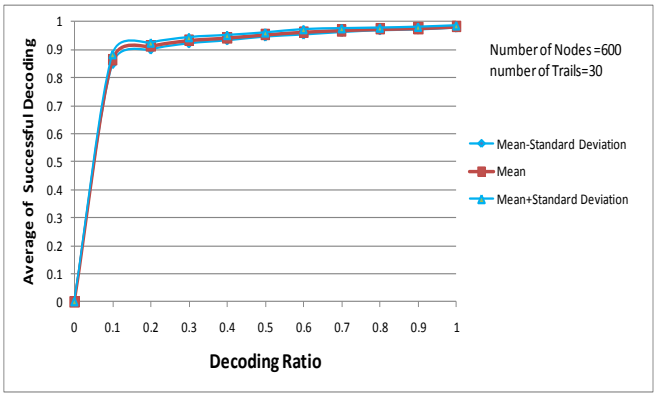


Figure.9.A WSN is randomly distributed in a field. The average of successful decoding ratio is shown for network size= 600 nodes and buffer size= 60 slots with the MDSA algorithm

VII. Conclusion and Future Work

Modified distributed storage algorithm (MDSA) for wireless sensor networks is proposed in this paper to enhance data recovering for WSNs .This is achieved by disseminate sensed data throughout the networkusing flooding and unicasting to provide controlled-redundancy of data through the network.

The results and performance evaluation demonstrated that MDSAalgorithmis better on average of the successful decoding percentage according the same evaluationmetrics with the previous algorithm (DSA-I).

MDSA has shown better performance through increasing the network size.MDSA works to reduce unnecessary power consumption and memory usage when possible.Our future work will include practical and implementation aspects of this algorithm.

ACKNOWLEDGMENT

I want to thank first my advisor,Prof.Dr.Mostafa-Sami M. Mostafa for his supervision, encouragement, and his guidance. Also, I would like to thank my supervisor Dr.HossamEldeenMoustafaShamardan, for his valuable advices and guiding me during this research work.

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam and E. Cayirci, "A survey on sensor networks," IEEE Communications Magazine, August 2002.

[2] S A Aly, H Darwish, M Youssef and M. Zidan,"Distributed Flooding-based storage algorithms for large-scale wireless sensor networks," In Proc. IEEE International Conference on Communication, Dresden, Germany, June 13-17, 2009.

[3] Philip A. Chou and Yunnan Wu," Network Coding for the Internet and Wireless Networks", Microsoft Research One Microsoft Way, Redmond, WA, 98052, June 2007.

[4] Zhenzhou Tang, Hongyu Wang, Qian Hu, and Long Hai, "How Network Coding Benefits Converge-Cast in Wireless Sensor Networks", In Proc. IEEE Vehicular Technology Conference (VTC 2012 Fall),2012.

[5] A. G. Dimakis, P B Godfrey, M. Wainwright and K. Ramchandran, "The Benefits of Network coding for peer-to-peer storage," In Proc. Of 26th IEEE Infocom, Anchorage, AK, USA, May 6-12, 2007.

[6] Y. Lin, B. Liang, and B. Li. "Data persistence in large-scale sensor networks with decentralized fountain codes", In Proc. Of the 26th IEEE INFOCOM07, Anchorage, AK, May 6-12, 2007.

[7] A. Kamra, V. Misra, J. Feldman and D. Rubenstein, "Growth codes: Maximizing sensor network data persistence", In Proc. 2006 conference on Applications, Technologies, Architec- tures, and Protocols for Computer Communications, pp 255-266, Pisa, Italy, 2006.

[8] Salah A. Aly, Zhenning Kong and Emina Soljanin, "Fountain Codes Based Distributed Storage Algorithms for Large-scale Wireless Sensor Networks", International Conference on Information Processing in Sensor Networks. 2008.

AUTHORS PROFILE

Mohamed Labib Borham is currently a Teaching Assistan, Faculty of Computer Science, Modern Sciences and Arts University (MSA, , 6 Octobe, , Egypt. He is a Masters Student at the Computer Science Department, Faculty of Computers and Information, Helwan University, Cairo, Egypt. He received her B.Sc. In Computer Science from Suez Canal University ,Ismalia. His current research interests include Wireless Sensor Networks Algorithms and applications,Artificial intelligence.

Mostafa-Sami M. Mostafa is currently a Professor of computer science, Faculty of Computers and Information, Helwan University, Cairo, Egypt. He worked as an Ex-Dean of faculty of Computers and Information Technology, MUST, Cairo. He worked also as an Ex-Dean of student affairs and Ex-Head of Computer Science Department, Faculty of Computers and Information, Helwan University, Cairo, Egypt. He is a Computer Engineer graduated 1967, MTC, Cairo, Egypt. He received his MSC 1977 and his PhD 1980 from University of Paul Sabatier, Toulouse, France. His research activities are in Software Engineering and Computer Networking. He is awarded supervising more than 80 Masters of Sc. And 18 PhDs in system modeling and design, software testing, middleware system development, real-time systems, computer graphics and animation, virtual reality, network security, wireless sensor networks and biomedical engineering.

Hossam Eldeen Moustafa Shamardan is currently a lecturer of Information Technology, Faculty of Computers and Information, Helwan University, Cairo, Egypt. He is an Electronics Engineer graduated 1991, Faculty of Engineering, Cairo University, Egypt. He received his MSC 1996 and his PhD 2005 from the University of Cairo, Giza , Egypt. His research activities are in Information security, Image processing, wireless sensor networks, and Computer Networking.

A Distributed Method to Localization for Mobile Sensor Networks based on the convex hull

Yassine SABRI
STIC Laboratory
Chouaib Doukkali University, B.P: 20
El Jadida MOROCCO

Najib EL KAMOUN
STIC Laboratory
Chouaib Doukkali University, B.P: 20
El Jadida MOROCCO

***Abstract*— There has been recently a trend of exploiting the heterogeneity in WSNs and the mobility of either the sensor nodes or the sink nodes to facilitate data dissemination in WSNs. Recently, there has been much focus on mobile sensor networks, and we have even seen the development of small-profile sensing devices that are able to control their own movement. Although it has been shown that mobility alleviates several issues relating to sensor network coverage and connectivity, many challenges remain. Among these, the need for position estimation is perhaps the most important. Not only is localization required to understand sensor data in a spatial context, but also for navigation, a key feature of mobile sensors. This paper concerns the localization problem in the case where all nodes in the network (anchors and others sensors) are mobile. We propose the technique following the capabilities of nodes. Thus, each node obtains either an exact position or an approximate position with the knowledge of the maximal error born. Also, we adapt the periods where nodes invoke their localization. Simulation results show the performances of our method in term of accuracy and determinate the technique the more adapted related to the network configurations.**

Keywords- wireless sensor network (WSN); Mobility; Localization; scalability.

I. INTRODUCTION

A wireless sensor network is composed of a large number of small and inexpensive smart sensors for many monitoring, surveillance and control applications. Each sensor makes its own local observation. All active sensors in the network coordinate to provide a global view of the monitored area. It is anticipated that such a network can be used in unattended environments or hostile physical locations. Applications include habitat monitoring [1][2], infrastructure surveillance [3], target tracking in tactical environments [4], etc. Almost all these applications require sensors to be aware of their physical locations. For example, the physical positions should be reported together with the corresponding observations in wildlife tracking, weather monitoring, location-based authentication, etc [5][6][7]. Location information can also be used to facilitate network functions such as packet routing [8][9] and collaborative signal processing [10], in which the complexity and processing overhead can be substantially reduced. Further, each node can be uniquely identified with its position, thus exempting the difficulty of assigning a unique ID before deployment [11]. However, many challenges exist in designing effective and efficient sensor self-positioning schemes for sensor networks. First, a localization algorithm must scale well to large sensor networks. Further, the location discovery scheme should not aggravate the communication and computation overheads of the network, since low-cost sensors have limited resource budgets such as battery supply, CPU, memory, etc. What's more, the localization scheme should not raise the construction cost of sensor nodes. Finally, the positioning scheme should be robust enough to provide high precision even under noisy environments.

This paper deals with the problem of localization in wireless sensor networks when sensors are mobile. There are three scenarios of mobility: sensors and anchors are mobile; sensors are mobile and anchors are static; sensors are static and anchors are mobile. For the last case, some methods have been proposed [12], [13]. In these methods, mobile anchors can be robots, humains, or other, equipped GPS which are used in order to locate others static sensors. In this paper, we present a new method to resolve the localization problem in the complex scenario where nodes and anchors are mobile. However, this method can be used for the two others cases of mobility. Three schemes are proposed following the capabilities of sensors. Sensors can be equipped with techniques like ToA/TdoA (Time of arrival / Time difference of arrival) or RSSI (Received Signal Strength Indicator) allowing computing distance between a pair of neighbor sensors. They may also be equipped with AoA (Angle of arrival) technique allowing computing angle between a pair of neighbor sensors. Finally, sensors may be equipped by none of these techniques. Our method determines an exact position for a sensor when it has at least two anchors in its neighborhood. Otherwise, it gives an approximate position and can compute in this case the generated maximal error. The localization problem with mobile sensors introduces a new problem: in fact, the energy of sensors being weak, each node cannot compute continually its localization in order to maintain accuracy position during its move.

Therefore, the question is: when a node must evoke the calculation of its position? In [14], authors compare three methods Static Fixed Rate (SFR), Dynamic Velocity Monotonic (DVM) and Mobility Aware Dead Reckoning Driven (MADRD). These methods define periods during which sensors should invoke their localizations. However, the authors assume that when a node invokes its position it obtains an exact localization (e.g. all sensors are equipped with GPS). These methods are explained in section II. However, when only a small number of sensors are anchors, the problem is not addressed. In this paper, we consider this case of network.

When a node invokes its localization it does not always obtains its exact position: either it obtains an approximate position or it cannot locate itself. To overcome this problem, our method defines the periods when a node has to invoke its location. Finally, through simulations, we analyze performances of our three techniques.

The rest of the paper is organized as follows: In Section 2, we summarize related work on localization algorithms. In Section 3, introduces basic notions for this problem. In Section 4 and 5, we present our new localization algorithm. In Section 6, we evaluate the proposed scheme through comprehensive simulation studies. We conclude the paper in Section 6.

II. Related Work And Background

First, The popular Global Positioning System (GPS) [15] localization system may not be a practical solution for outdoor sensor networks. It is infeasible to install GPS on each sensor due to cost, form factors, power consumption and antenna requirements. Further, GPS requires direct Light-Of-Sight (LoS) communication, which renders it unfeasible for many outdoor application environments. Therefore in the past several years, extensive research has been directed to designing GPS-less location discovery schemes [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34], [7], [35], [36] .These positioning algorithms differ in their assumptions on network deployment, device capabilities, node mobility, signal propagation, error requirement, etc. Thus, they can be classified differently.

For example, some methods are designed for static sensor networks, where sensors remain stationary after deployment, while others are for dynamic sensor networks where sensors and beacons are mobile [40]. These localization schemes can also be classified as centralized [37], [33], [39], where all computations are performed by a central point (e.g., the base station), or distributed, where sensors estimate their positions independently of each other. Centralized methods have poor scalability and are thus infeasible for large sensor networks. In this section, we will focus on distributed location discovery schemes for stationary sensor networks, which can be further classified as beacon-based and beacon-less depending on whether or not beacons are used, or classified as range-based and range-free according to the type of knowledge used in position estimation.

The majority of current location detection systems assume the existence of beacons, whose positions are known through GPS receivers or manual configuration. A typical sensor first measures the distances or angles from it to several beacons, and then obtains position estimation through techniques such as triangulation, trilateration, multilateration, etc. Based on the coverage capabilities of beacons, these localization systems can be further classified as systems with long-range beacons or systems with short-range beacons. Systems with long-range base stations have a fixed set of powerful beacons, whose transmission range can cover the entire network. Usually these base stations are manually deployed, are time-synchronized, and are equipped with special instruments such as directional antennas. In systems with shortrange beacons [33], [34], [7], [35], a small number of sensors with known positions are randomly deployed amongst other ordinary sensors. Some of them rely on transmitting both RF and ultrasound signals at the same time, where the RF is used for time-synchronizing the sender and the receiver.

Range-based localization relies on the availability of point-to-point distance or angle information. The distance/angle can be obtained by measuring Arrival (ToA), Time-Difference-of-Arrival (TDOA), Received-Signal-Strength-Indicator (RSSI), and Angle-of-Arrival (AOA), etc. The range-based localization may produce fine-grained resolution, but have strict requirements on signal measurements and time synchronization. ToA measures the signal arrival times and calculates distances based on transmission times and speeds. GPS [16] is the most popular ToA-based localization system. By precisely synchronizing with a satellite's clock, GPS computes node position based on signal propagation time. Compared to ToA, TDoA has an advantage as the former's processing delays and non-LOS propagation can introduce larger errors [38]. Ref. [7] proposes a TDoA based scheme (AHLos) that requires base stations to transmit both ultrasound and RF signals simultaneously.

The RF signal is used for synchronization purposes. A sensor first measures the difference of the arrival times between the two signals, then determines the range to the base station. Finally, multilateration is applied to combine range estimates and generate location data. RSSI computes distance based on transmitted and received power levels, and a radio propagation model. RSSI is mainly used with RF signals [16], but the range estimation can be inaccurate due to multipath fading in outdoor environments [7]. AoA-based methods first measure the angle at which a signal arrives at a base station or a sensor, and then estimates the position using triangulation.

The calculation is quite simple, but AoA techniques require special antenna and may not perform well due to omnidirectional multipath reflections. Further, the signals can be difficult to measure accurately if a sensor is surrounded by scattering objects [12]. To the best of our knowledge, there is no other method, in the open literature that deals with this case. Some papers have been proposed in the case where anchors are mobiles and others sensors are static.

For example, [12], [13] uses robots or humans, which can be considered as anchors, move in the network and help others nodes to obtains their positions. When sensors are mobile, it is not reasonable that each sensor invokes its localization technique in order to locate itself continually, due to constraint of energy. A first work in [14] proposes three methods SFR (Static Fixed Rate), DVM (Dynamic Velocity Monotonic), MADRD (Mobility Aware Dead Reckoning Driven) to determinate periods where a node invokes its localization technique. But, it assumes that a node obtains its exact position when it invokes its localization (e.g. sensors are equipped with GPS). The following sub-sections explain these three methods.

A mobile sensor changes its position with time. A simple strategy for finding its position is the use of standard localization methods at any time. But if the position of the sensor is required frequently, this method is very costly. Tilak et. al tried to reduce the frequency of localizations for finding

the position of sensors. They proposed techniques: Static Fixed Rate (SFR), dynamic Velocity Monotonic (DVM) and Mobility Aware Dead Reckoning Driven (MADRD). SFR calls a classical localization operation periodically with a fixed time interval. To respond a query from the base station, a sensor sends its position obtained from the last localization. When a sensor remains still or moves fast, in both cases, the reported position suffers a large error. In DVM, localization is called adaptively with the mobility of the sensors. The time interval for the next call for localization is calculated as the time required to traverse the threshold distance (a distance, traversed by the sensor, location estimation assumed to be error prone) with the velocity of the sensor between last two points in the sequence of localization calls. In case of high mobility, a sensor calls localization frequently. If a sensor suddenly moves with very high speed from rest, error in the estimated location becomes very high. In MADRD, the velocity is calculated from the information obtained from last two localized points.

The predictor estimates the position with this velocity and communicates to the query sender. At the localization point, the localized position is reported to the query sender and the distance error is calculated as the distance between the predicted position and reported position.

III. PRELIMINARY

Before In this paper, we focus on mobile sensor network. Moreover, we assume that all the sensors have identical transmission radius r. however, it is easy to adapt our method with sensors having different transmission radius. We represent a wireless sensor network as a graph $G(V,E)$ where V is the set of n nodes representing sensors and E is the set of m edges representing communication links. If two nodes u, v are neighbors, then they are linked and the distance between u and v is smaller than r. We assume also that some anchors have a priori knowledge of their own positions with respect to some global coordinate system (GPS) (black nodes in figures). We consider scenarios where nodes and anchors are mobile.

For example, in a military context, soldiers can be equipped with sensors and tanks with anchors. Soldiers use tank positions in order to obtain their positions. Finally, we should take into account functionalities of each sensors: for example, methods like RSSI or ToA/TDoA and AoA described in previous section. A wireless sensor networks is represented as a bidirectional graph $G(V,E)$ where V is the set of n nodes representing sensors and E is the set of m edges representing communication links. If two nodes $u,v \in V$ are neighbors, then they are linked that means distance between u and v is smaller than r. The set of neighbors for a node $u \in V$ is noted $N(u)$.Anchor nodes have knowledge of their location through some other means, such as GPS or simply explicit programming. The set of anchors is noted Λ. The set of neighbor anchors for a node u is noted $N_\Lambda(u)(N_\Lambda(u) = N(u) \cap \Lambda)$ and the set of non-neighbor anchors is noted $\overline{N_\Lambda}(u)(\overline{N_\Lambda}(u) = \Lambda / N_\Lambda(u))$. Note that all identical nodes (anchors or others nodes) have the same capabilities (energy, processing, communication, ...). The coordinate of a position of node u is noted (x_u, y_u). P is the set of all possible positions in a network. Our method construct the convex hull of a point cloud S_u for each node u ,this convex hull is noted $conv(\mathsf{S}_u)$. The localization modules (eg, GPS or Galileo) are expensive and consumers of energy,for this our method seeks to use the least possible anchors with the Nodes can use technology measures distances as ToA, RSSI, AoA. So, when it receives a signal from a transmitter, a node deduces that it is located on the circle centered on the transmitter. The exact distance between two nodes u and v is noted d_{uv}. Two neighbor nodes u, v know d_{uv} (via ToA, ...). The estimated distance is noted $\hat{d}_{uv}$. The following section explains how to obtain these estimated distance. the set of circles built from the knowledge of anchor neighbors is noted C_{N_Λ} ,the set of circles built from the knowledge of non-anchor neighbors is noted $\mathsf{C}_{\overline{N_\Lambda}}$. ε is the distance between the estimated position $(x_{u_{estm_i}}, y_{u_{estm_i}})$ of the sensor u and the summit furthest from convex hull $Conv(\mathsf{S})$. Let d_{err} being the distance between the estimated position of a node and its real position, representing the position error. The node knows that d_{err}. By using a predefined threshold, if $d_{err} \leq threshold$ then the node has an estimation close to its real position. In this case the node becomes an estimated anchor and broadcasts its position .

IV. LOCALIZATION TECHNIQUE

A. Localization Algorithm Based on the convex hull

The Initially, each anchor broadcasts its position. A node can therefore be deduced the distance between each of the anchors We use the technique SumDist (Savvides et al., 2002) for estimating distances adding the distances between separated sensor nodes of an anchor.Upon receiving the position of a anchor, a node considers the following cases:

• If it receives directly the position of the anchor, he deduces they are neighbors and therefore it located on the circle centered at the anchor or radius of a circle is r.

• If it receives the position by an intermediate node, it concluded that it is not neighbor of the anchor and therefore it is not inside the circle of radius r centered in anchor . So,when a node u receives a position of an anchor A, it estimates the distance to this anchor with Sum-Dist and draws one or two circles. In fact, if $(A \in N_\Lambda(u))$, u knows d_{Au} and deduces that it is on the circle C_{Au} of radius equals to d_{Au} and centered in A. If $(A \notin N_\Lambda(u))$ then u knows that it is not inside the circle of center A and radius r otherwise A and u would be neighbors. Moreover, u knows the estimated distance to A, $\hat{d}_{Au}$ deduced by Sum-Dist. By triangular

inequality, $\hat{d}_{Au} \le d_{Au}$. u applies this technique to each received anchor position. So, u is inside the circle $C_{\overline{Au}}$ of center A and radius $\hat{d}_{Au}$. Thus, the intersection of circles defines a cloud of points S_u . the center of gravity of the convex hull of this cloud $conv(S_u)$ represents the estimated position of u .

To summarize, for each node $u \in V/\Lambda$, the envelope obtained as follow :

Initialization of the algorithm:

$$S_0 = P \quad (1)$$

$$C_{N_\Lambda(u)_0} = C_{\overline{N_\Lambda}(u)_0} = \{\varnothing\} \quad (2)$$

When a node u receives a message controle P from anchor node a_i neighbor:

If $a_i \in N_\Lambda(u)$:

The circle centered at a_i and of radius $d^2_{ua_i}$:

$$C_{ua_i} = \{(x_i, y_i) \in P \mid (x_i - x_a)^2 + (y_i - y_a)^2 = d_{ua_i}\} \quad (3)$$

Construction of intersection points of a circle C_{ua_i} with the old circles $C_{(u)_{i-1}}$,keeping only the points inside a circle centered at u and of radius $\varepsilon_{u_{i-1}}$:

$$W_{u_i} = \{(x_i, y_i) \in (C_{(u)_{i-1}} \cap C_{ua_i}) \mid (x_i - x_{u_{estm_i}})^2 + (y_i - y_{u_{estm_i}})^2 \le \varepsilon^2_{u_{i-1}}\} \quad (4)$$

Cleaning the old cloud of points S_{i-1} ,keeping only the points inside a circle centered at u and of radius d_{ua_i} :

$$Z_{u_i} = \{(x_i, y_i) \in S_{i-1} \mid (x_i - x_a)^2 + (y_i - y_a)^2 \le d^2_{ua_i}\} \quad (5)$$

New cloud of points S_i :

$$S_i = Z_{u_i} \cup W_{u_i}, i \ge 3 \quad (6)$$

The circle C_{ua_i} joins the old circles $C_{N_\Lambda(u)_{i-1}}$:

$$C_{N_\Lambda(u)_i} = C_{ua_i} \cup C_{N_\Lambda(u)_{i-1}} \quad (7)$$

Same effect occurs when a node u receives a message controle P from anchor node a_i not neighbor :

if $a_i \notin \overline{N_\Lambda}(u)$:

$$C_{\overline{ua_i}} = \{(x_i, y_i) \in P \mid (x_i - x_a)^2 + (y_i - y_a)^2 = \hat{d}^2_{ua_i}\} \quad (8)$$

$$W_{u_i} = \{(x_i, y_i) \in (C_{(u)_{i-1}} \cap C_{\overline{ua_i}}) \mid (x_i - x_{u_{estm_i}})^2 + (y_i - y_{u_{estm_i}})^2 \le \varepsilon^2_{u_i - 1}\} \quad (9)$$

$$Z_{u_i} = \{(x_i, y_i) \in S_{i-1} \mid r^2 \le (x_i - x_a)^2 + (y_i - y_a)^2 \le \hat{d}^2_{ua_i}\} \quad (10)$$

$$S_i = Z_{u_i} \cup W_{u_i}, i \ge 3 \quad (11)$$

$$C_{\overline{N_\Lambda}(u)_i} = C_{\overline{ua_i}} \cup C_{\overline{N_\Lambda}(u)_{i-1}} \quad (12)$$

$$C_{(u)_i} = C_{N_\Lambda(u)_i} \cup C_{\overline{N_\Lambda}(u)_i} \quad (13)$$

The end for each node we will have a set of points S_u of the cloud:

$$S_u = \{p_1, p_2, p_3, \cdots, p_n\} \quad (14)$$

Calculate the convex hull S_u based on Jarvis' March:

$$conv(S_u) = \{\sum_{n=0}^{n} \alpha_i p_i \mid \alpha_i \ge 0, \sum_i \alpha_i = 1\} \quad (15)$$

The new estimation error ε_{u_i} :

$$\varepsilon_{u_i} = \max_{p \in conv(S)} d(p, u_{estm_i}) \quad (16)$$

The main design of the Slsnj, which is a simple finite state machine. As shown in figure 1, a node running Slsnj is in one of four states at any time: (i) Sensor not estimated, (ii) Sensor estimated, (iii) estimated Anchor,and (iv) improve the accuracy. Transitions between the states are triggered by events. After the Slsnj protocol is initiated, the node enters the Sensor not estimated state,Whenever the node receives a broadcasting ProbePacket packet, the node enters the Sensor not estimated state and uses this packet to estimate its postion,after this stage of estimation the node switches to another state is depending on the value of the estimation error found,if *espilon<threshold* the node enters in estimated Anchor state else it enters in Sensor estimated state .In the latter two states a node is still waiting of probpacket packet from anchor or estimated Anchor nodes to enter in improve the accuracy state and improve its accuracy.when there will be no more ProbePacket, the node switches to the state final and considered as estimated with an error of precison. An example is illustrated in figure 2. X receives positions of anchors A, B and C . It estimates distances $\hat{d}_{AX}$, $\hat{d}_{BX}$, $\hat{d}_{CX}$ with Sum-Dist. Since all an- chors are not neighbors of X then X is not inside circles centered respectively in A, B, C with a radius equals to r but it is inside circles with radius equal to $\hat{d}_{AX}$, $\hat{d}_{BX}$, $\hat{d}_{CX}$. The intersection of these circles defines the cloud points S_X for a node X . X computes the center of gravity of the convex hull $conv(S_X)$ of this cloud and estimates its position in $G2$.

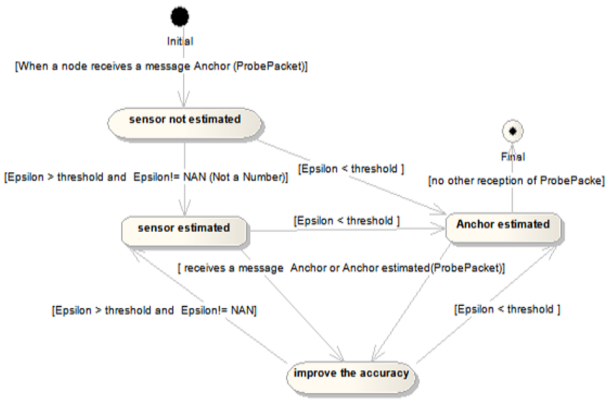


Figure 1: State machine diagram for Sensor node not estimated

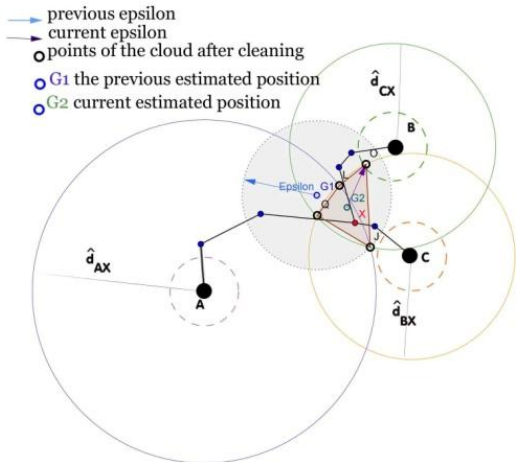


Figure 2: Example of estimating the position for X at the reception of localization information from C anchor

B. Pseudo-code

The pseudo-code for the Slsnj is shown in figure 3. Each anchor exact (equipped with GPS or Galileo) or estimated broadcasts its position through the control message P,and depending on number of hops traveled by the packet P we check its validity,if the number of hops is less than a certain threshold called *ThresholdHopcount* it is considered confirmed otherwise a packet is rejected.after we apply our method as described previously .

C. P roperties

Our localization technique meets three very important properties who have a significant impact on its performance:

• First, a node knows if its estimated position is close to its real position. Let ε be the distance between the center of gravity and the point, in the zone, furthest away from the center of gravity. Let d_{err} being the distance between the estimated position of a node and its real position, representing the position error. The node knows that $d_{err} \leq \varepsilon$.

Algorithm IV.1: When a node n receives a message controle (P)

```
global List_Anchors, comment: structure contains all the received anchor
List_Pts_Intersection, comment: structure contains the set of points of the cloud S_u
MyCercl_epsilon, comment: circle centered at the estimated position and of radius ε_u
local rcved_Anchor, comment: received anchor
cercl_epsilon_tmp, comment: temporary variable to calculate the estimated position and the estimation error
if (P.gethopcount() ≤ Threshold_hopcount)
    P.sethopcount(P.hopcount+1)
    send (ProbePacket (P))
    if (Isneighbour (P))
        then { rcved_Anchor.exterior_crcl ← cercl (P.getX(), P.getY(), distance (myself, Pneighbour))
               rcved_Anchor.interior_crcl ← cercl (P.getX(), P.getY(), 0)
        else { rcved_Anchor.exterior_crcl ← cercl (P.getX(), P.getY(), P.Dist + P.epsilon)
               rcved_Anchor.interior_crcl ← cercl (P.getX(), P.getY(), radius - P.epsilon)
    if (List_Anchors.Isempty())
        then {List_Anchors.add (rcved_Anchor)
        then { List_Pts_Intersection ← Intersection (List_Anchors, rcved_Anchor)
               List_Pts_Intersection ← Clean_Failed_Pts (MyCercl_epsilon, rcved_Anchor)
               List_Pts_Intersection ← Graham_Scan (List_Pts_Intersection)
               cercl_epsilon_tmp.centr ← Centre_of_gravity (List_Pts_Intersection)
               List_Anchors.add (rcved_Anchor)
               cercl_epsilon_tmp.radius ← calc_epsilon (cercl_epsilon_tmp.centr , List_Pts_Intersection)
        else { if (cercl_epsilon_tmp.radius ≤ MyCercl_epsilon.radius)
                   then { MyCercl_epsilon.radius ← cercl_epsilon_tmp.radius
                          MyCercl_epsilon.radius ← cercl_epsilon_tmp.centr
                          if (cercl_epsilon_tmp.radius ≤ Threshold_AnchorEstimated)
                              then { setImAnchorEstimated (true)
                                     send (ProbePacket (myID, BrodcastID, 2, MyCercl_epsilon, pktSequence+1))
```

Figure 3: Description of algorithm Slsnj

By using a predefined *threshold* , if $\varepsilon \leq threshold$ then the node has an estimation close to its real position. In this case the node becomes an estimated anchor and broadcasts its position and its ε . When a node applies the approximation technique with an estimated anchor radius, it takes into account ε .Consider a sensor X calculating its position with an estimated anchor A. If they are neighbors, X trace two circles (belongs to C_{N_Λ}) centered in A of radius $d_{AX} \pm \varepsilon$ and deduce that it is between these two circles. If they are not neighbors, X deduces that it is not inside the circles centered at A of radius $r - \varepsilon$ and belongs to a circle of radius $d_{AX} + \varepsilon$,the definitions (4),(6),(9) and (11) become:

si $a_i \in N_\Lambda(u)$:

$$C_{ua_i} = \{(x_i, y_i) \in P \mid (x_i - x_a)^2 + (y_i - y_a)^2 = (d_{ua_i} \pm \varepsilon_{u_i})^2\}$$
$$Z_{u_i} = \{(x_i, y_i) \in S_{i-1} \mid (x_i - x_a)^2 + (y_i - y_a)^2 \leq (d_{ua_i} \pm \varepsilon_{u_i})^2\} \quad (18)$$

si $a_i \notin \overline{N_\Lambda}(u)$:

$$C_{\overline{ua_i}} = \{(x_i, y_i) \in P \mid (x_i - x_a)^2 + (y_i - y_a)^2 = (\hat{d}_{ua_i} \pm \varepsilon)^2\} \quad (19)$$
$$Z_{u_i} = \{(x_i, y_i) \in S_{i-1} \mid (r - \varepsilon)^2 \leq (x_i - x_a)^2 + (y_i - y_a)^2 \leq (\hat{d}_{ua_i} + \varepsilon)^2\} \quad (20)$$

• Second, a node can detect if some informations are wrong. This case is illustrated in esxpresion W_{u_i} .With its bound error ε , nodes reject the cloud points that are outside of circle centered at its estimated position and of radius ε .for example, when a node u detects a point of its cloud S_u it outside in the circle centered at u of radius ε will not take it into account . This property is defined by the expression W_{u_i} .

• Third, convex hull algorithms as Graham scan [?] and Jarvis march [?] allowed us to calculate the convex hull $conv(S)$ a cloud of points with a very optimum complexity , of order $O(n\log(n))$ with n the number of points of the cloud, which allowed us to reduce consumption of CPU time (and therefore energy), but also allowed us to optimize particularly the consumption of memory storage ,focusing not on global interpretation of the network as in an algorithm of type Grid-scan, but only on points of the cloud. The improvement made allowed us to retain the properties functional Our localization technique despite the change in network size, and efficiently localize the nodes (continuously) and with a certain level of quality in different scales.

D. Structure of the control message exchanged

Our approach Requires the exchange of Specific Information. For this, a specific control message is designed. The fields in this message , called ProbePacket, exchanged during the execution of the localization algorithm are shown in Table 1 ,tow possible values for the packet subject are used in the algorithm:Anchor,Anchor estimated. Note that when a node broadcasts or sends a message in a wireless network, all

nodes in its scope communication receive this message. The validation of a control message is limited by a threshold of validation ,called *Threshold_hopcount*.

SrcId	global identifier ID of the transmitter node of the packet
DestId	global identifier ID the receiving node
NodeType	type of the transmitter node of the packet (Anchor, Anchor estimated)
MyCercl_epsilon	circle centered at the estimated position or exact position of the transmitter , and of radius ϵ (0 for transmitter Anchor)
Hopcount	the number of hops between the anchor transmitter and the sensor receiver
pktSequence	to avoid the redundantly reception of the same message
Dist	the sum of the distances intermediate between the anchor transmitter and the receiver sensor

Figure 4: Fields of the message ProbePacket

V. ADAPTATION OF DVM AND MADRD

DVM and MADRD determine periods when a node has to invoke its localization technique, related to mobility of nodes. It is necessary to adapt these two techniques in order to take into account accuracy of localization. SFR is not concerned by this problem because its period of time is constant. In these techniques, when a node is moving fast, localization will be carried out more often and conversely. But if a node is located with important error, it is necessary to invoke localization technique more often. Therefore, if node is located with high accuracy,methods DVM and MADRD do not need any change but if node obtains an approximate position then protocols DVM and MADRD have to take into account the error ε .Let t be the time returned by DVM or MADRD and t' the time returned by our method when ε is taken into account.If $\varepsilon = 0$ (ie. the position is exact)then $t = t'$ and if $\varepsilon \geq r$ (ie.the position is bad then $t' = 0$.

Between these two values , t' varies linearly $t' = t - \varepsilon \frac{t}{r}$ .Thus,If ε represents an important error,then periods during which a node should invokes its localization will be short and conversely if ε is a small error. *Perturbation of predictions in* MADRD : In MADRD nodes calculate their positions related to predictions.A node computes its position related to its previous position. conversely if a small error. Perturbation of predictions in MADRD : In MADRD nodes calculate their positions related to predictions.A node computes its position related to its previous position.

VI. SIMULATIONS

A. Environnement de simulation

Experiments were built upon the J-Sim simulator [9] dedicated to WSN simulations. It is a compositional, component based simulation environment. It is built upon the concept of autonomous component programming model. J-Sim is developed entirely in Java. The signal attenuation due to obstacles or other factors (e.g. use of unidirectional antennas) is simulated in J-Sim. Therefore, the vicinity of a node in terms of transmission range is not necessarily spherical. Note that there several simulators in the literature such as GlomoSim[41] , OMNET++[42] , OPNET[43] , NS-2[44] . The MAC layer is considered perfect and the transmission of messages are without loss in our simulation.

In order to allow easy comparison between different scenarios, range errors as well as errors on estimated positions are normalized to the radio range. For example, of position error means a distance of half the range of the radio between the real and estimated positions. The percentage of range errors is noted .

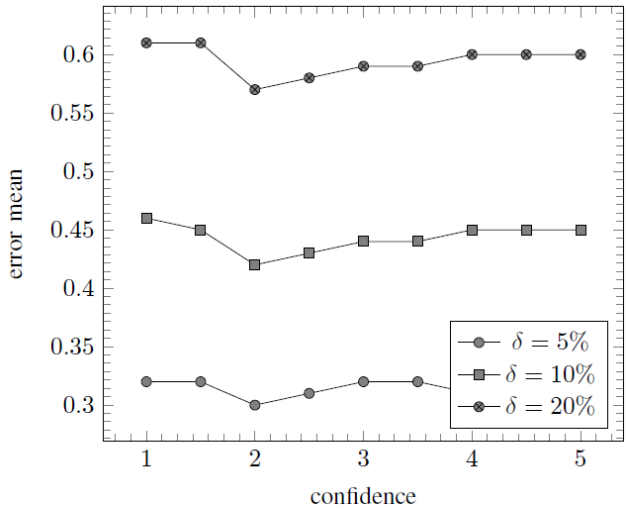


Figure 5: impact of threshold confidence

B. Results

Globally, the positions determined by a localization algorithm represent a geometrical layout of the physical positions of the sensors. This layout must be compared to the ground truth, or known layout of the sensors. It is important therefore that not only the error between the estimated and real position of each node is minimized, but also that the geometric layout determined by the algorithm matches well the original geometric layout. In order to have a unified approach for evaluate the accuracy of our technique and a solid frame for analysis of the scalability, we propose to use two metrics.

• **MAE(Mean Absolute Error):** The simplest way to describe localization performance is to determine the residual error between the estimated and actual node positions for every node in the network, sum them and average the result. Broxton et al in [45] do this using the mean absolute error metric (MAE), which, for each of n nodes in the network, calculates the residual between the estimated nodes and actual coordinates.

$$MAE = \frac{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 - (y_i - \hat{y}_i)^2}{n} \quad (21)$$

with (x_i, y_i) the real position and $(\hat{x}_i, \hat{y}_i)$ the estimated postilions .

• **GDE (Global Distance Error):** As discussed briefly at the start , it is important for the accuracy metric to reflect not only the positional error in terms of distance, but also in terms of the geometry of the network localization result. GDE in

[46] takes the RMS error over the network of n nodes and normalises it using the constant R. In Ahmed et als context, R represents average radio range, meaning the localization results are represented as a percentage of the average distance nodes can communicate over.

$$GDE = \frac{1}{r}\sqrt{\frac{\sum_{i=1}^{n}\sum_{j=i+1}^{n}(\frac{\hat{d}_{ij}-d_{ij}}{d_{ij}})^2}{n(n-1)/2}} \quad (22)$$

with $\hat{d}_{ij}$ The estimated distance between i and j and d_{ij} The actual distance between i and j.

This section analyses the performances of our three methods related to the techniques SFR, DVM and MADRD. Mobility model: The mobility model used in this paper is the random waypoint model [47]. It is the classical model used in the mobile network. In this model, velocities of nodes vary and a node can stop its move. Each node picks a random location and starts moving to it. As soon as the node reaches the destination, it picks another destination randomly and moves toward it. Our simulation uses the BonnMotion tool to generate the various scenarios of mobility where velocity and trajectory deviation of nodes vary. Each scenario runs during 90 seconds.

Simulation model: In our simulations, all messages are delivered. For easier comparison between different scenarios, range errors as well as estimations of position errors are normalized to the radio range. This technique is classical in the literature and allows comparisons with others methods. For example, of position error means a distance equal to half of the radio range between the real and estimated positions. In our scenarios, we use nodes in a square of . The transmission range of nodes is equal to . Among nodes, we randomly select anchors with representing a density of anchors in the square from to . Also, we consider measure errors of , and respectively. Analyse: In our method, it is possible that a node does not obtain an estimated position when it does not contain anchors in its neighborhood. This case depends on the anchors density. Therefore, if our simulations consider only the position average error rate of sensors, performances of our three techniques would not be shown due to this case. As a consequence, our results focus on the time during which a node is located with a position error lower than for MAE metric and for GDE metric. After this time, nodes are considered that they are badly positioned. For our analysis, we perform tests. For each scenario, we take into account the mean and we represent on graphs the confidence interval. Here, there is of chance that the real values belong to this interval.

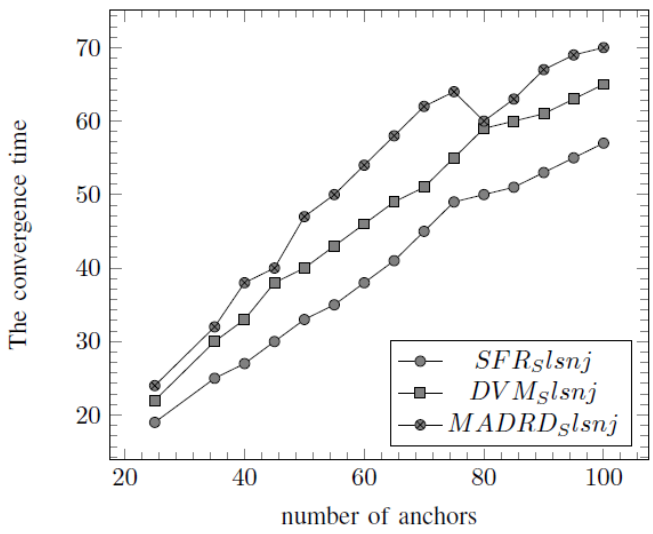


Figure 6: GDE Performances of SFR, DVM and MADRD with errors equal to 5%

1) Without measure errors : In this section, we consider the ideal case where measure errors are equal to. The figures 7 and 9 show simulations with SFR, DVM and MADRD for GDE and MAE. These curves represent the time during which a node is located with a position error lower than for MAE for GDE. For example, in figures 5, when the network contains 70 anchors, a node is located with an error lower than during: *41.01s*. Without surprise, accuracy of positions is based on the capability of nodes to calculate distances. MADRD provides better results than DVM and then SFR.

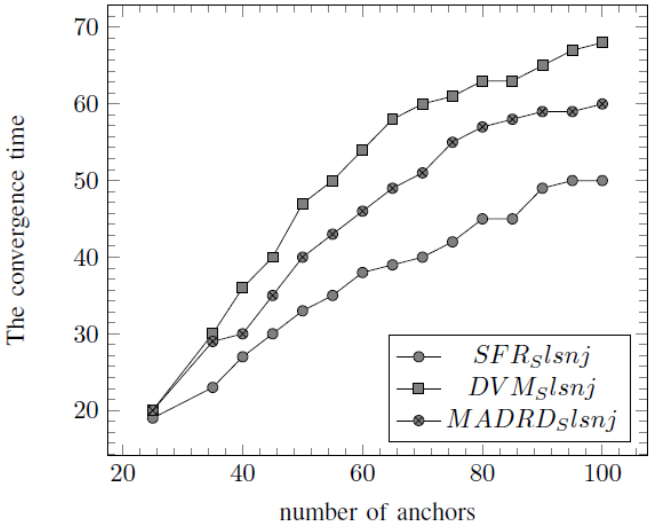


Figure 7: GDE Performances of SFR, DVM and MADRD without measure Errors

2) Measure errors equal to 5% *:* In this section, we introduce measure errors equal to 5%. Figure 6 and 8 shows that results obtained when nodes can calculate distances, is not influenced too much by measure errors in SFR, DVM and MADRD for MAE and GDE. To conclude, DVM provides better results than MADRD and then SFR.

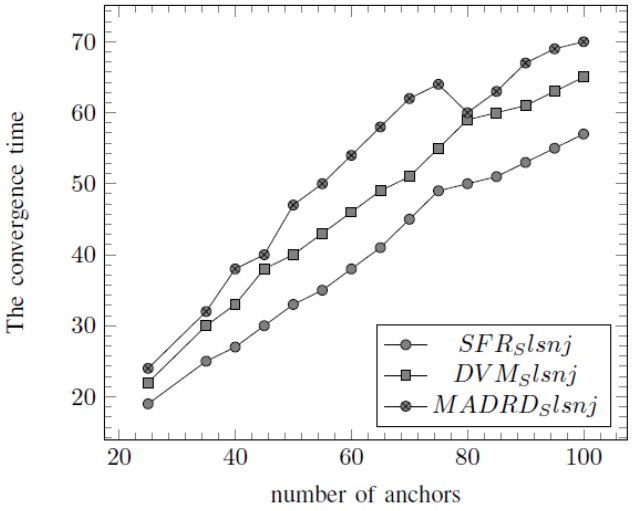


Figure 8: MAE Performances of SFR, DVM and MADRD with errors equal to 5%

3) Conclusions of simulations : In These simulations show the performances of our method and show how to adapt SFR, DVM and MADRD, related to the network environment in order to provide good results. We note the impact of measure errors in MADRD since it is efficient only if it uses accurate positions. MADRD provides good results in a network environment without measure errors, but when we introduce errors, DVM is the best. Finally, phenomenons seen in an environment with measure errors equal to 5 errors equal to 10.

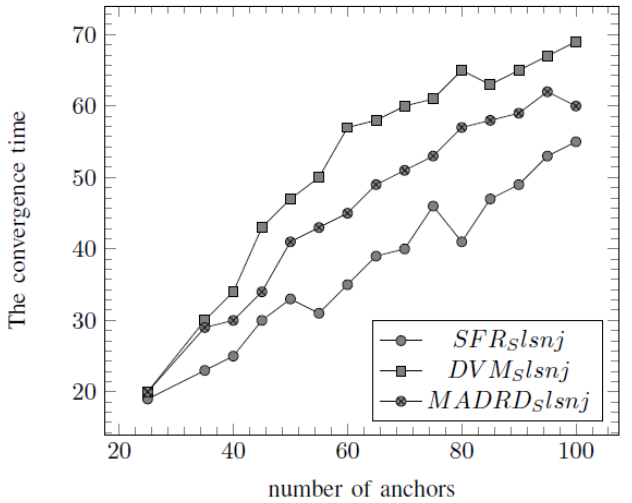


Figure 9: MAE Performances of SFR, DVM and MADRD without measure errors

VII. Conclusions

This paper proposes the method for the localization problem when anchors and others sensors are mobile. this method take into account capabilities of nodes: nodes which can calculate either distances with their neighbors . Moreover, in order to answer to question when a node should invoke its position? related to network environment and capabilities of nodes, we adapted techniques SFR, DVM, MADRD, proposed in [14]. Our simulations show the performances of our method and determinate the technique the more adapted related to the network configurations.

References

[1] Cerpa, J. Elson, M. Hamilton, J. Zhao, D. Estrin, and communications technology," in Workshop on Data communication in Latin America and the Caribbean, ser. SIGCOMM LA '01. New York, NY, USA: ACM, 2001, pp. 20–41.

[2] A. Mainwaring, D. Culler, J. Polastre, R. Szewczyk, and J. Anderson,"Wireless sensor networks for habitat monitoring," in Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications, ser. WSNA '02. New York, NY, USA: ACM, 2002, pp. 88–97.

[3] T. van Dam and K. Langendoen, "An adaptive energy-efficient mac protocol for wireless sensor networks," in Proceedings of the 1st international conference on Embedded networked sensor systems, ser. SenSys '03. New York, NY, USA: ACM, 2003, pp. 171–180. [Online]. Available : http ://doi.acm.org/10.1145/958491.958512

[4] Q. Fang, F. Zhao, and L. Guibas, "Lightweight sensing and communication protocols for target enumeration and aggregation," in Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing, ser. MobiHoc '03. New York, NY, USA : ACM, 2003, pp. 165–176.

[5] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, "Directed diffusion for wireless sensor networking," IEEE/ACM Trans. Netw., vol. 11, no. 1, pp. 2–16, Feb. 2003. [Online]. Available : http ://dx.doi.org/10.1109/TNET.2002.808417

[6] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong, "Tag: a tiny aggregation service for ad-hoc sensor networks," SIGOPS Oper. Syst. Rev., vol. 36, no. SI, pp. 131–146, Dec. 2002.

[7] A. Savvides, C.-C. Han, and M. B. Strivastava, "Dynamic fine-grained localization in ad-hoc networks of sensors," in Proceedings of the 7th annual international conference on Mobile computing and networking, ser. MobiCom '01. New York, NY, USA : ACM, 2001, pp. 166–179.

[8] S. De, C. Qiao, and H. Wu, "Meshed multipath routing with selective forwarding: an efficient strategy in wireless sensor networks," Computer Networks, vol. 43, no. 4, pp. 481–497, 2003.

[9] J. Li, J. Jannotti, D. S. J. De Couto, D. R. Karger, and R. Morris, "A scalable location service for geographic ad hoc L. Girod, "Habitat monitoring: application driver for wireless For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation routing," in Proceedings of the 6th annual international conference on Mobile computing and networking, ser. MobiCom '00. New York, NY, USA: ACM, 2000, pp. 120–130.

[10] J. Heidemann and N. Bulusu, "Using geospatial information in sensor networks," 2001.

[11] C. Schurgers, G. Kulkarni, and M. B. Srivastava, "Distributed ondemand address assignment in wireless sensor networks," IEEE Trans. Parallel Distrib. Syst., vol. 13, no. 10, pp. 1056–1065, Oct. 2002.

[12] N. B. Priyantha, H. Balakrishnan, E. Demaine, and S. Teller, "Mobile-Assisted Localization in Wireless Sensor Networks," in IEEE INFOCOM, Miami, FL, March 2005.

[13] N. Bulusu, J. Heidemann, and D. Estrin, "Adaptive beacon placement," 2001, pp. 489–498.

[14] S. Tilak, V. Kolar, N. B. Abu-Ghazaleh, and K.-D. Kang, "Dynamic localization protocols for mobile sensor networks," CoRR, vol. cs.NI/0408042, 2004.

[15] B. Hofmann-Wellenhof, H. Lichtenegger, and J. Collins, Global Positioning System: Theory and Practice, 5th ed. Springer, Feb. 2001.

[16] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol. 2, pp. 775–784 vol.2, 2000.

[17] N. Bulusu, J. Heidemann, and D. Estrin, "Gps-less low cost outdoor localization for very small devices," IEEE Personal Communications Magazine, vol. 7, no. 5, pp. 28–34, October 2000.

[18] "Adaptive beacon placement," in Proceedings of the The 21st International Conference on Distributed Computing Systems, ser. ICDCS '01. Washington, DC, USA: IEEE Computer Society, 2001, pp. 489–.

[19] S. Capkun and J.-P. Hubaux, "Secure positioning of wireless devices with application to sensor networks," in INFOCOM, 2005, pp. 1917– 1928.

[20] X. Cheng, A. Thaeler, G. Xue, and D. Chen, "Tps: A time-based positioning scheme for outdoor wireless sensor networks," in IEEE INFOCOM, 2004, pp. 2685–2696.

[21] L. Fang and W. Du, "A beacon-less location discovery scheme for wireless sensor networks," in In Proceedings of IEEE INFOCOM, 2005, pp. 13–17.

[22] L. Girod and D. Estrin, "Robust range estimation using acoustic and multimodal sensing," vol. 3, 2001.

[23] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in Proceedings of the 9th annual international conference on Mobile computing and networking, ser. MobiCom '03. New York, NY, USA : ACM, 2003, pp. 81–95.

[24] Jeffrey, "SpotON : An Indoor 3D Location Sensing Technology Based on RF Signal Strength

[25] L. Hu and D. Evans, "Localization for mobile sensor networks," in Proceedings of the 10th annual international conference on Mobile computing and networking, ser. MobiCom '04. New York, NY, USA : ACM, 2004, pp. 45–57.

[26] K. Langendoen and N. Reijers, "Distributed localization in wireless sensor networks : a quantitative comparison," Comput. Netw., vol. 43, no. 4, pp. 499–518, Nov. 2003.

[27] L. Lazos and R. Poovendran, "Serloc : Robust localization for wireless sensor networks," ACM Trans. Sen. Netw., vol. 1, no. 1, pp. 73–100, Aug. 2005.

[28] Isaac Amundson and Xenofon D. Koutsoukos. 2009. A survey on localization for mobile wireless sensor networks. In Proceedings of the 2nd international conference on Mobile entity localization and tracking in GPS-less environments (MELT'09), Richard Fuller and Xenofon D. Koutsoukos (Eds.). Springer-Verlag, Berlin, Heidelberg, 235-254.

[29] A. Nasipuri and K. Li, "A directionality based location discovery scheme for wireless sensor networks," in Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications, ser. WSNA '02. New York, NY, USA: ACM, 2002, pp. 105–111.

[30] D. Niculescu and B. Nath, "Ad hoc positioning system (aps," in In GLOBECOM, 2001, pp. 2926–2931.

[31] ——,"DV Based Positioning in Ad Hoc Networks," Telecommunication Systems, vol. 22, no. 1, pp. 267–280, Jan. 2003.

[32] A. Rao, S. Ratnasamy, C. Papadimitriou, S. Shenker, and I. Stoica, "Geographic routing without location information," in Proceedings of the 9th annual international conference on Mobile computing and networking, ser. MobiCom '03. New York, NY, USA: ACM, 2003, pp. 96–108.

[33] N. Sastry, U. Shankar, and D. Wagner, "Secure verification of location claims," in Proceedings of the 2nd ACM workshop on Wireless security, ser. WiSe '03. New York, NY, USA: ACM, 2003, pp. 1–10.

[34] C. Savarese, J. M. Rabaey, and K. Langendoen, "Robust positioning algorithms for distributed ad-hoc wireless sensor networks," in Proceedings of the General Track of the annual conference on USENIX Annual Technical Conference, ser. ATEC '02. Berkeley, CA, USA: USENIX Association, 2002, pp. 317–327.

[35] A. Savvides, H. Park, and M. B. Srivastava, "The n-hop multilateration primitive for node localization problems," Mob. Netw. Appl., vol. 8, no. 4, pp. 443–451, Aug. 2003.

[36] Y. Shang, W. Ruml, Y. Zhang, and M. P. J. Fromherz, "Localization from mere connectivity," in Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing, ser. MobiHoc '03. New York, NY, USA; ACM, 2003, pp. 201–212.

[37] L. Doherty, K. S. J. Pister, and L. El Ghaoui, "Convex position estimation in wireless sensor networks," in INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE, vol. 3, 2001, pp. 1655–1663.

[38] Xuemin Shen, Jon W. Mark, and Jun Ye. 2002. Mobile Location Estimation in CDMA Cellular Networks by Using Fuzzy Logic. Wirel. Pers. Commun. 22, 1 (July 2002), 57-70. DOI=10.1023/A:1016025802932 http://dx.doi.org/10.1023/A:1016025802932.

[39] R. Graham, "An efficient algorithm for determining the convex hull of a finite planar set," Information Processing Letters, pp. 132–133.

[40] R. A. Jarvis, "On the identification of the convex hull of a finite set of points in the plane," Inf. Process. Lett., vol. 2, no. 1, pp. 18–21, 1973.

[41] "About glomosim," http://pcl.cs.ucla.edu/projects/glomosim/, cited July 2011.

[42] "Omnet++ community site," http://www.omnetpp.org/, cited July 2011.

[43] "Opnet technologies," http://www.opnet.com/, cited July 2010.

[44] "The network simulator," http://www.isi.edu/nsnam/ns/, cited July 2010.

[45] M. Broxton, J. Lifton, and J. A. Paradiso, "Localization on the pushpin computing sensor network using spectral graph drawing and mesh relaxation," SIGMOBILE Mob. Comput. Commun. Rev., vol. 10, pp. 1–12, January 2006.

[46] A. A. Ahmed, H. Shi, and Y. Shang, "Sharp : A new approach to relative localization in wireless sensor networks," in Proceedings of the Second International Workshop on Wireless Ad Hoc Networking - Volume 09, ser. ICDCSW '05. Washington, DC, USA: IEEE Computer Society, 2005, pp. 892–898.

[47] T. Camp, J. Boleng, and V. Davies, "A survey of mobility models for ad hoc network research," WIRELESS COMMUNICATIONS & MOBILE COMPUTING (WCMC): SPECIAL ISSUE ON MOBILE AD HOC NETWORKING : RESEARCH, TRENDS AND APPLICATIONS, vol. 2, pp. 483–502, 2002.

A Strategy to Improve The Usage of ICT in The Kingdom of Saudi Arabia Primary School

Gafar Almalki
(School of Computer Science, Engineering and Mathematics): Flinders University
Adelaide, Australia

Neville Williams
(School of Computer Science, Engineering and Mathematics): Flinders University
Adelaide, Australia

***Abstract*— Integration of ICT in education is a complex idea that requires practical interpretation to get significant outcomes. As a developing country, the Kingdom of Saudi Arabia (the KSA) does not have a proper technological infrastructure as developed countries. Efficient strategies are vital in improving the application of ICT in the KSA's primary schools effectively. Improving the usage of ICT in the KSA primary schools achievement entails integrating ICT into classroom. However, some barriers that prevent a successful ICT implementation in the primary school are still present. This paper proposes several strategies to trounce the challenges. Several recommendations for ICT integration in primary school applicable in the case of the KSA are also necessary. These strategies are executable at school and national scale.**

Keywords- ICT; primary school; barrier; strategy.

I. INTRODUCTION

Information and communication technologies (ICT) are indispensable tools for any organization to stay competitive and play their role efficiently and effectively. A significance gain is achievable by organizations, which adopt and implement ICT successfully [1]. In the education sector, ICT is also a potential tool to deliver high quality education and prepare students for the information era [2]. When computers, as part of ICT, integrate into the educational system, several expectations will arise such as a reduction of teachers' efforts to deliver knowledge, achievement of better visualization, and encouragement of students' motivation [Marcuse-Hass and Kromenholtz cited in 3]. However, integration of ICT in education is a complex idea, which practical interpretation interpreted is a significant outcome [4].

As other developing countries, the Kingdom of Saudi Arabia (the KSA) has no proper technological infrastructures like developed countries. Therefore, efficient strategies are crucial to improve the usage of ICT in the KSA primary schools effectively. Integrating ICT into classroom improves the usage of technology in the KSA primary schools. However, some barriers, which prevent a successful ICT implementation in the primary schools, still prevail. An analysis to understand how teachers and students utilize ICT and impacts of teaching and learning process in the classrooms is extremely valuable to construction of strategies to overcome the problems. An intensive review of the ICT usage in the schools is essential. It aims at identifying and categorizing useful strategies, which are appropriate and applicable in the KSA.

The aim of the study is to develop a set of strategies to improve the usage of information and communication technologies (ICT) in the Kingdom of Saudi Arabia primary school. The categories of the usage of ICT are 'supportive ICT use' and 'classroom ICT use'. Supportive ICT use refers to the application of ICT for practical, educational support such as school administration, teaching management and administration and preparation of worksheets for student assignment. Classroom ICT use is the implementation of ICT in teaching and learning process in the classroom and science labs such as use of computers for demonstration and visualization, drilling ,and practice activities [Tondeur et al. cited in 5].

II. REVIEW TECHNOLOGICAL USAGES IN PRIMARY SCHOOL

A. Types of Educational Computer Use

Most of The ICT research studies explore the conditions that can sustain its integration into schools. In this context, many researchers present frameworks or models demonstrating conditions that can have an influence on ICT integration into teaching and learning methods. These frameworks are mostly on two basic methods: 'qualitative research methods' [for example, 6] and 'quantitative research methods' [for instance, 7]. All the aforementioned frameworks have one thing in common that ICT integration is described from a holistic point of view under the influence of conditions situated on different levels (pupils, teachers, schools, and policy makers) [7].

Based on an empirical study involving a large segment of teachers, Tondeur et al. [cited in 5] describes two main categories of ICT use by teachers: supportive ICT use and classroom ICT use. Supportive ICT refers to the use of ICT for practical as well as organizational teaching responsibilities such as, student administration, teaching management and administration, preparing worksheets, developing evaluation activities presuming students' learning progress in a given situation.

The next category, 'classroom ICT use' focuses on supporting and heightening the actual teaching and learning processes, for instance, how computers are applicable in

demonstration purposes, drilling and practice activities. It also entails representation of intricate knowledge elements, deliberations, cooperation, and project work. Furthermore, Wozneyet et al. [cited in 5] finds that supportive use of ICT is the most decisive factor to appraise the classroom use of ICT.

B. Characteristics of ICT Usage

Tondeu, Valcke, & van Braak [8] explore the individual characteristics of teachers and school environment that relates to various types of computer usage in primary schools. There was a survey in Flanders primary schools whereby administration of questionnaires relates to (ICT) coordinators from the same schools to collect auxiliary information in accordance with the cultural and contextual school characteristics.

Strenuous efforts existed, in the quest to determine the effects of teachers and relative school features in the investigation. As the study reveals that cultural school characteristics, constituting the schools' readiness to change as well as the availability of an ICT based school policy plan, links to the use of computers as an educational instrument. The cultural school also links to the implementation of ICT with consideration to the fundamental computer skills. On the contrary, no cultural school characteristic is likely to have links independently, with the use of computers as an information instrument. A multidimensional approach can bring a more conductive and meaningful insight to the characteristics affecting the use of computers.

III. Analysis

The Ministry of Education, the Ministry of Higher Education and the General Organization for Technical Education and Vocational Training have joint responsibility to manage education system in the KSA. Kindergarten is for children aged 3–5 years, but it is not a condition for enrollment in primary educations.

The first grade of primary education is for children aged six years. They must pass the final examination at the end of Grade 6 and obtain the Elementary Education Certificate in order to move to intermediate education. The students study at intermediate education for three years. They also must study for three years at secondary education, which acts as the final stage of general education. Students can go to higher education subsequent to going through secondary education [9]. Table 1 shows the number of students and teachers at each level in 2007 [10].

Number of Students and Teachers in 2007

Education Level	Students		Teachers	
	Male	Female	Male	Female
Primary	1,255,117	1,187,365	107,227	110,328
Intermediate	609,300	535,248	54,034	54,031
Secondary	541,849	471,225	41,108	46,715

ICT have a potential to change the ways of teaching. However, to achieve satisfactory results, some consideration on barriers that hinder the implementation gives the best way to start. Balanskat, Blamire and Kefala [cited in 11] state that although teachers acknowledge the importance of ICT to improve the learning process, some problems continue to arise during the process of adopting these technologies. This section describes several problems on the implementation of ICT in several countries, which may reflect in the KSA.

An undertaken analysis represents an understanding of how teachers and students utilize ICT and what the impacts for teaching and learning process in the classrooms. Researchers have used several categories to classify barriers relating to implementation of ICT on school. Hew & Brush [6] identified the barriers into several groups, namely: (a) resources, (b) institution, (c) subject culture, (d) attitudes and beliefs, (e) knowledge and skills, and (f) assessment. Bingimlas [11] divided the barriers into (a) teacher level and (b) school level. Mohamed, Abuzaid & Benladen [12] regarded the barriers as cultural and technical constraints. This analysis concentrates on teacher factor, school/institution factor and extrinsic factor.

A. Teacher Factor

Teacher-related serves as the most significant predictor for technology adoption. Teachers are critical key to successful adoption of ICT while teachers' professional development needs to focus on technology skills and support for pedagogical change to embed ICT' [13]. Teachers should be at the center of ICT adoption projects. Marcinkiewicz [c,ted in 5] stressed that the full integration of ICT in education will remain a remote target to achieve, unless the gap between teachers perceived impediments and the use of computers occurs caution. The three categories related to the teachers are:

1) Lack of teacher confidence

Several references reported that inadequacy in teachers' poise prevents them to use ICT on class. Becta [cited in 11] found that teachers were afraid to use ICT in the classroom as they feel having limited knowledge in the area of ICT. Lack of confidence will also reduce teachers' enthusiasm in using ICT as a useful tool in the teaching process [11]. A study conducted in 2005 by Jamieson-Proctor et al. [14] indicates that 73% of female teacher in Queensland State, are less confident to use ICT for teaching and learning process in classrooms. This number was significantly greater than their male counterparts were.

2) Lack of teacher competence

Inadequate teacher competence is a barrier, which influences negatively teachers' confidence. William et. al. [cited in 6] found that 10% more of elementary school teachers in Scottish school have not enough skills on databases and spreadsheet. In Australia, Newhouse [cited in 11] found that many teachers were not enthusiastic to integrate ICT with teaching activities since they do not have enough knowledge and skill about this technology. In United States, Snoeyink and Ertmer [6] also found that lack of computer skill was becoming a barrier for teachers to integrate ICT into the teaching process on the classrooms.

In Malaysia, lack of teacher expertise on ICT is the main constraint of promoting ICT in schools [15]. A study by Cavas et al. [16] in Turkey found that Turkish primary science teachers' positive approach to ICT, received impact from their computer ownership at home and their computer experience.

3) Negative attitudes

Teachers' participation is significant since it influences the success of ICT implementation in the schools [7]. Schoepp [cited in 11] states that although some teachers have enough knowledge and skill on ICT, they did not intend to use ICT because of the belief that there are not appropriate rewards for using it. Furthermore, some teachers believed that there is no relationship between ICT and an effective teaching [6]. For example, a study in Australia revealed that most of the teachers believed a better learning could not occur by using computer [Newhouse cited in 6]. In addition, Jamieson-Proctor et al. [14] found that there is significant resistance from teachers in Queensland State to use ICT as required by the curriculum.

The teachers who have an obsession and sheer commitment for boosting their learning processes can more likely incorporate technology in their teaching. This view is also consistent with findings by Sang et al. [5] that amongst variables of the internal teachers, ICT incentive appears to be the ICT classroom exploit strongest predictor. In addition, attitudes of teachers towards the application of ICT in education relates to ICT motivation as part of the predictor of ICT classroom use.

B. School/Institution Factor

1) Lack of time

Using ICT on the classrooms needs a reliable preparation, which may consume much time. Teachers need many hours to search suitable resources on the Web, prepare photos and videos for multimedia presentation. If the schools do not reduce teachers' teaching time as compensation, it is difficult to expect the teachers to use ICT on their classroom [6]. Furthermore, availability of computers in laboratories does not guarantee the teachers to access it if they should compete with the other teachers for laboratory time [Zhao, Pugh, Sheldon, and Byers cited in 6]. In Saudi Arabia, Al-Alwani [cited in 11] found that teachers do not want to use ICT as they have a busy schedule.

2) Lack of efficient training

Lack of teacher competence and skill is the main barrier of ICT classroom integration. A good training for teacher may be a solution for the problem. Therefore, lack of effective training for the teacher is a significant barrier for the implementation of ICT on classroom [19]. Gomes [cited in 11] states that insufficient training in digital literacy, pedagogic and didactic training becomes barriers for teachers in applying ICT in the classrooms.

3) Lack of local technical support

Neyland [18] states that provision of ICT and technical support are first order factors affecting technology integration'. It is common that the teachers face technical problem when working in technology-integrated-classroom. On this classroom, there are various kinds of technological resources such as computers, LCD projector, printers, CD-ROMs, whiteboard and other multimedia device.

It is hard to encourage teachers to use ICT without immediate support from skilled staffs. This is when the teachers have difficulties in operating the technological resources [Lim et al. cited in 6]. Indeed, short of technical assistance is significant constraints for using ICT [Pelgrum cited in 11]. In New Zealand, a study by Education Review Office of New Zealand revealed that technical support was a significant factor for successful use of schools' e-learning packages [18]. In Saudi Arabia, teachers were not in agreement with the idea of applying ICT without technical support [Almohaissin cited in 11].

4) Leadership barrier

A good example from the leaders (school principal) is a significant factor for the successful integration of ICT in school. Stuart et al. [20] state that the leadership behaviors of senior management contribute pivotally in determining the success or failure of an ICT implementation'. Fox and Henri [cited in 6] reported that most of the teachers in Hong Kong had no interest in using ICT in the classroom. This results from the feeling support deficiency from the principal who did not understand ICT. In Victoria, Australia, the successful implementation of ICT into the curriculum requires the commitment of the school leaders to arrange suitable training for the teachers [21].

C. Extrinsic Factor

1) Local culture

One of the considerable advantages of using ICT is the possibility to deliver information over the internet, which performs e learning. However, unsuccessful adoption of e learning in developing countries resulted from direct copy from developed countries that have significantly different cultures [Wurm 2008 cited in 12].

A few more studies have explored how these factors influence in a direct and/or indirect way the levels of ICT integration in classrooms. Although the question about ICT integration in education has global significance, yet the cultural variables have, should also be part of the consideration. For instance, cultural differences identified through comparing the Chinese and Flemish teacher perspectives in the use of ICT in teaching and learning methodologies [Zhu et al. cited in 5].

According to their findings, Chinese teachers, in particular, express more doubts about the constructivist principles underlying many ICT applications inclusive but not limited to collaboration, independent learning, and self-directed learning. These differences are noticeable with the ideas of Chinese teachers as compared to the Flemish teachers concerning teacher–student and student–student interactions. This links to disparity in the cultural dimensions of both societies e.g., power distance, collaboration, and competition significantly. Chinese teachers nevertheless put a larger emphasis on those nuances.

Implementation of ICT continues through the provision of access to the Internet. There are many negative views about contents of the Internet such as extreme political beliefs, pornography, and strange religions [Postrel cited in 22]. Although all of these factors may be not considered ethically abhorrent' in some western societies, it cannot be accepted in the KSA culture which bases on the Islamic religion [22]. Indeed, after a long national debate about the social risks of the Internet, the internet introduction in the KSA in 1999 was to support e-learning [Wurm 2008 cited in 12]. Modification of the example strategies from other countries is necessary for ICT adoption in the KSA.

2) Lack of funding support

Integrating ICT into the school program requires a high cost for the supply and maintenance of this technology [23]. Sufficient funds are one of the keys to successful implementation of ICT in the schools [15]. Without sufficient funds, the schools cannot provide software and hardware such as computers and its peripherals. Hew & Brush [6] state that without adequacy of these technological resources, it was difficult to encourage teachers to use ICT in the classrooms. Indeed, a sufficient fund is one of an important factor to ensure sustainability ICT integration in the primary school.

3) Lack of appropriate planning

A successful ICT integration requires a good planning. Technology alone was not sufficient to integrate ICT in the classrooms. A comprehensive plan for introducing ICT in the society was also required [Mool cited in 23]. Somekh [13] states that 'the complexities of educational innovations require a holistic strategy capable of building change in social practices informed by the practical power of theoretical knowledge'. A good comprehensive planning is essential to solve simultaneously problems such as lack of appropriate planning, lack of funding support, coping different local culture and lack of effective training.

IV. Discussion And Initiatives

Various references propose a varied solution for the implementation of ICT in schools. This section examines solutions adaptable to KSA.

A. *Personal Development for Teachers*

Some recent researches reveal that irrespective of teachers' earnest inclination, profound interest in knowing the potential of ICT, their practical use of ICT is comparatively low, and it generally focuses on a limited range of applications. On the other hand, another international survey on teachers' perceived impediments in using ICT reveals three main factors i.e. their lack of resources, their lack of knowledge and skillfulness and instructive complexities to incorporate technology in instruction. However, the competency and compatibility of teachers' self-assurance in their expertise, is a key to understand their eagerness to integrate technology in their instruction. Zhao and Cziko [cited in 25] explored three conditions that can help motivate teachers in using ICT to carry out their pedagogical responsibilities:

(1) Teachers should have faith that the use of technology would enhance the effectiveness of their goals, which is unachievable through ordinary means,

(2) They should also believe that the use of technology would not hamper the other high-level goals they want to accomplish, and

(3) They should believe that they have full command on available resources and have enough potential for effectively utilizing ICT. Cox [cited in 26] states that a positive perception of students and teachers to the value of ICT will encourage them to use ICT. Therefore, understanding the teachers' perception toward ICT will help the decision maker (government or school principal) to make plans on how the teachers will adopt ICT in their teaching activities.

Professional development for teachers was considered as a critical factor in the successful integration of ICT in schools [Way & Webb cited in 18]. Teachers should have 'technology-supported-pedagogy knowledge' which is required when they plan to integrate ICT in classrooms [Hughes cited in 6]. By this knowledge, teachers may ask students to write their assignments in a word processor document rather than submitting a hand-written paper. The teachers may also give art assignments drawn on graphic software or a 2D/3D modeling software.

Training teachers enable teaching in 'student-centered way' and develop 'individual learning program' for their students. They also encourage the students to work independently on their own computer [23]. A school in Sidney conducted a regular workshop to increase teacher confidence in ICT. The workshop involved local experts and consultants to ensure a sustainability [23]. Furthermore, Jones [cited in 26] states that the barriers that hamper the teacher from using ICT can be eliminated through two ways, namely, individual level enablers and whole school level enablers. At individual level, the teachers have full access to personal PCs or laptops with good quality hardware and software, access to good educational resources, and appropriate training. At whole school level, adequate technical support, effective timetabling, support from senior staff or principal and adequate equipment, such as whiteboards in classroom are key factors.

To sum up, basic and general ICT training for all primary school teachers is required and offered at the national level by the government of the KSA. More advanced and specific ICT training should be at the school level and with its subjects determined by each school requirements.

B. *Training for Students*

Availability of ICT resources will not be useful if the students have no skill to use it in their learning process. Therefore, the school should provide basic training for the students. In Malaysia, there was program called 'Pintar', which delivered basic ICT training for students from low-income families. The training had the aim of increasing the motivation to learn in students [15]. Indeed, as training for teachers is given, suitable ICT training for students is also required.

C. *Institution/School Support*

Support from institution/school is required to perform 'supportive environment' which include encouragement to use ICT, training, and in providing technical support staff [28]. All ICT components, hardware and software, should be well maintained, with technical support available when teachers have difficulties to operate it [23]. Hiring special staff, that has the responsibility to support ICT implementation in the primary schools, may be a good solution.

A well planned timetable may provide the teachers more time to prepare teaching material which use ICT [28]. By reducing teachers' workload, they will have more time for developing learning program with their colleagues, trying new method to teach in ICT environment, and thinking about better pedagogical practices [23]. Schools should also facilitate work groups, which enable the teachers to work together in producing learning materials. This effort will save teachers time, while increase their productivity.

As a leader, the schools principals' commitment to improve the usage of ICT will support the successful integration of ICT in the school. This support can be in the form of improvement of curriculum and establishing technology committee, encouraging teachers to improve their technological skills, providing appropriate resources and continuous monitoring. The technology committee main task is the development of plans and strategies to ensure that ICT integration will work well [23]. Furthermore, from questionnaires to teachers on online learning tool integration in schools, Neyland [18] concluded that 'local leadership, including the level of school support and commitment to innovation, was seen to be more important than broader systemic level strategies'.

The effort by schools in integrating ICT in science laboratories will assist students in getting intriguing visualizations of nature proceedings. This effort will be inspirational to students in familiarizing with ICT alongside developing their learning capabilities [29].

D. *Community and Government Support*

When schools have adequate finances, ICT access sustainability will be a guarantee. Government financing is a top prerequisite for the provision of appropriate software and hardware to the schools. Support from finances raised by the association of parents allows for the continuity and access achievement in the access of the ICT[28].

Involving commercial ventures is also a significant step. In India, there was software provision by Pearson Education that combined the ELearning tools, administration tools, management of homework tools, and management of teacher's resources. There has been a successful running of pilot projects in 125 schools. The tools possibly will assist the schools with the implementation of ICT along with saving their worthy time[30].

1) Example of State / National Scale Project in Developed Countries

Technology in itself was not satisfactory for ICT integration in classrooms. A detailed plan for the introduction of ICT in the social order was elementary [Mool cited in 23]. The government ought to develop an integrated plan and for an ICT implementation structure on a national scope.

In Australia, the NSW government undertook a 'Computer in School Plan' project, which involved updating computer hardware and software regularly, providing computers for use, and developing an enhancement curriculum [23]. In UK, the government launched City Learning Centers (CLCs) to integrate ICT into schools. This program has increased usage of ICT in lesson planning and teaching programs as the teachers have increased access to ICT.

There is a provision of various trainings on the computer and package software operations. As a result, the teachers became increasingly confident in operating computers in during their lessons [31]. Furthermore, on January 2002, the Welsh Assembly Government in UK incurred £9.9 million from grants to provide all the primary schools with interactive whiteboards (IWB), computers, and projectors. The IWB displays projected images and permits users to have power over the connected PC through a touch on the board. Even though, other teachers require convincing regarding the usefulness of IWBs, most believe the tool is extremely significant for the future of teaching [Kennewell & Morgan cited in 17].

In South Korea, the integration of ICT in education is a significant factor necessary for the attainment of a rapid economic growth and development. Since 1970, the government of South Korea has been providing considerable finances towards the integration of ICT in education. The use of ICT in education, prepared the South Korea population to a quick adaption of science and technology in the labor market [Kim cited in 32]. In 1990, the government of South Korea introduced an eight-year plan of offering trainings to teachers on the use of ICT. The plan was part of the government's master plan for education. The training was an immense success and became a basic pillar for the integration of ICT in the education system of South Korea [32].

2) Example State / National Scale Project in Developing Countries

Chile had Enlaces as its national program, which was useful for the implementation of ICT in schools alongside its integration in the curriculum. The objective of the program was to advance the quality of national education through the provision of a suitable ICT infrastructure, teachers' training, and implementation of digital resources. The project was a massive success as it was sustainable for both the rural and urban areas [33].

To reduce the risk of ICT-marginalization in the Arab World, Dutta et al. [34] suggests, 'devise a clear and comprehensive ICT development plan, supported by the highest political constituencies' and 'incorporate ICT skills and knowledge into the educational system'. Egypt acts as the first example in the Arab World. Egypt's Ministry of Communications and Information Technology centers its attention and energy towards expansion, utilization of the human resources required for the development of telecommunications, and IT sectors.

The government introduced an exclusive professional training program that will produce 5,000-trained professionals in IT per-annum. The ministry also plans to open a National Information Technology Institute (NITI) from which it will endeavor to patronize all the training programs necessary for the progression of IT skills across the country. The ministry plans to set up technical universities in the country and encourage the sending of young professionals abroad for further training. The ministry also considers taking the necessary measures needed for the advancement of the ICT based curriculum in Egypt [34].

Likewise, in Kuwait, the Ministry of Education recently launched a provision plan for broadband internet access to about 300 government schools through the issuance of tenders to large companies. The companies are to install basic infrastructures and connection facilities among many others. There is a complete involvement of the private sector in the implementation process, and expansion of the governmental policies towards ICT. This is through the provision of several discounts and concessions to private schools by decreasing the monthly internet charges and leasing of connections to the American and English schools in Kuwait [34]. In the U.A.E., the Ministry of Education, and Youth in collaboration with some foreign donors launched the project of Smart Schools with the purpose of encouraging and enhancing internet usage in schools. The mega project is inclusive of numerous incentives such as free internet installation in the private, government schools, and a discounted usage fees. There are numerous school subscriptions in the U.A.E. to the services[34].

E. *Solving Cultural Problem*

Unsuccessful adoption of ICT strategies could be due to direct borrowing from the developed countries that have different cultures with the KSA. For example, a successful implementation of the ICT plan requires continuous provision of internet access. In avoiding cultural conflicts regarding internet access, Sait et al. [22] suggests that the government should have a policy/legislation that will control and standardize the suitable contents. The government of the KSA employs several efforts towards the blocking of websites, which provide unacceptable content in the KSA such as pornographic materials, online gambling and dating. The use of central proxy server(software and hardware system) proved effective and did not cause significant delay for the users [Al-Furaih cited in 22].

Education in the KSA applies a gender-segregation policy which prohibits contact among men and women students [12]. Therefore, ICT execution to sustain e-learning should be modified. Mohamed et al. [12] proposed a customized e-learning system which limited communication among students from different gender. Only learners and tutors could interact.

Using ICT in the classes passes as adopting Western culture by most of the teachers in the KSA [26]. Therefore, Oyaid (2009) suggested that teachers should ensure that using ICT is compatible with their values, faith, and beliefs. Moreover, acquiring this new technology will make them better teachers as well as a better Muslims. Indeed, this effort may change teachers' negative attitude regarding ICT integration in classrooms into positive attitude.

F. *Recommendation for Implementation in the KSA*

The researchers proposed the implementation of several recommendations in the KSA based on several strategies that other countries have adopted:

1) Create a Suitable ICT Infrastructure Environment

An ICT system consists of hardware, software, people who use them and communications technology such as the Internet. Increasing the usage of ICT in the KSA primary schools is achievable by integrating ICT into the classrooms. Typical ICT projects involve upgrades of equipment (hardware and software) or new installations of the network. However, purchasing complete equipments to support ICT integration require a huge amount of money. Therefore, an integrated planning to improve ICT infrastructure within a medium-term period (3 to 5 years) at each primary school is important. The infrastructure can enhance internet access in schools and facilitate e-learning. Moreover, part of the fund should facilitate construction of ICT labs in schools. This effort will help students to get an interesting visualization of the nature process and encourage them to be familiar with ICT and develop their learning skill.

2) Training for Ministry of Education Staff

The Ministry of Education should develop, manage and control the project of ICT integration in schools. Therefore, the ministry should have competence staff in this area to integrate sustainability concepts and policies into ICT planning, and design projects. The Ministry should also carry out intensive ICT trainings for its staff.

3) Developing ICT Training Program for Teachers in Schools

The ICT trainings should form part of teacher career development. The training materials include 'technology-supported pedagogy knowledge' that is required when they plan to integrate ICT in the classrooms. By this training, the teachers will increase their confidence with ICT and enhance their ability to teach in a 'student-centered way' and developing 'individual learning program' for their students. Specific ICT training for teachers is also required. The school requirements determine the subjects that the teacher will undertake during training.

V. CONCLUSION

This study found several barriers of the implementation of ICT in schools. These barriers are in three categories; teacher factor, school/institution factor and extrinsic factor. Several strategies for ICT integration in primary school, which the KSA may adopt, are proposed. At national scale, integrating ICT with the curriculum extends to introducing some subjects related with ICT to the primary schools. The government should come up with integrated planning to improve ICT infrastructure at each primary school.

The ministry should launch basic ICT trainings as part of teacher career development, with materials such as 'technology-supported-pedagogy knowledge', which is required when they plan to integrate ICT in the classrooms.

Trainings for school principals are also required to improve leadership capability and their commitment to successful integration ICT in the school. Finally, the government may support research to develop a low cost interactive whiteboard that is crucial to improve students' learning abilities and encourage them to participate in the learning process in the classroom.

At institution/school scale, the schools should provide specific ICT trainings for teachers with school requirements determining what subjects to take. Hiring staff for ICT technical support will enable a 'supportive environment', which encourages teachers to use ICT and ensure all of ICT components, hardware and software, are well maintained. Facilitating discussion groups is important as a medium of sharing ideas among teachers about integrating ICT into classes. Creating a good timetable and reducing teachers' workload may provide the teachers more time to prepare teaching materials which uses ICT, developing learning program with colleagues, trying new methods to teach in ICT environment, and thinking about better pedagogical practices. Finally, integrating ICT into science labs will help students to get an interesting visualization of the nature process and encourage them to be familiar with ICT and develop their learning skill.

Quantitative and qualitative studies to evaluate the effectiveness of each strategy are important. Examples of these studies are:

- A study to evaluate suitable ICT training methods for teachers, school principals and Ministry of Education staffs, which the Ministry of Education can implement in the KSA. The training should also encourage teachers to use ICT and become open minded to new technologies.
- There should be a study to produce appropriate learning materials based on ICT, which suits the culture in Saudi Arabia.
- There should be a study to produce an appropriate government policy facilitating quicker ICT in the KSA.
- There should be a study to develop curriculum, which involves ICT as an important part. The syllabuses should include ICT in the course.
- There should be a study to develop software in Arabic language as part of ICT integration in the classrooms. The software also could also support science labs.

The results of the studies are useful in developing more comprehensive and efficient framework for ICT integration in primary schools in the KSA.

REFERENCES

[1] S.S. Al-Gahtani, "Computer Technology Adoption in Saudi Arabia: Correlates of Perceived Innovation Attributes," Information Technology for Development, vol. 10, pp. 57-69, 2003.

[2] A. Abdulkafi, "Teachers' Attitudes toward Information and Communication Technologies: The Case of Syrian Efl Teachers," Computers & Education, vol. 47, no. 4, pp. 373-398, 2006.

[3] S. Romi and H. Zoabi, "The Influence of Computer Technology Learning Program on Attitudes toward Computers and Self-Esteem among Arab Dropout Youth," Educational Media International, vol. 40, no. 3-4, pp. 259-268, 2003.

[4] M. Robertson, I. Webb, and A. Fluck, "Seven Steps to Ict Integration". 2007, ACER Press: Camberwell, Vic.

[5] G. Sang, M. Valcke, J. van Braak, J. Tondeur, and C. Zhu, "Predicting Ict Integration into Classroom Teaching in Chinese Primary Schools: Exploring the Complex Interplay of Teacher-Related Variables," Journal of Computer Assisted Learning, vol. 27, no. 2, pp. 160-172, 2011.

[6] K.F. Hew and T. Brush, "Integrating Technology into K-12 Teaching and Learning: Current Knowledge Gaps and Recommendations for Future Research," Educational Technology Research and Development, vol. 55, pp. 223–252, 2007.

[7] R. Vanderlinde, S. Dexter, and J. van-Braak, "School-Based Ict Policy Plans in Primary Education: Elements, Typologies and Underlying Processes," British Journal of Educational Technology, vol. 42, no. 3, pp. 1-15, 2011.

[8] J. Tondeur, M. Valcke, and J. van Braak, "A Multidimensional Approach to Determinants of Computer Use in Primary Education: Teacher and School Characteristics," Journal of Computer Assisted Learning, vol. 24, pp. 494-506, 2008.

[9] UNESCO IBE (2007) Saudi Arabia. World Data on Education.

[10] [10] Kingdom of Saudi Arabia: Ministry of Economics and Planning (2008) Achievement of the Development Plans Facts and Figures Twenty-Fifth Issue 1390-1429h 1970-2008g.

[11] K.A. Bingimlas, "Barriers to the Successful Integration of Ict in Teaching and Learning Environments: A Review of the Literature," Eurasia Journal of Mathematics, Science and Technology Education, vol. 5, no. 3, pp. 235-245, 2009.

[12] A.H. Mohamed, R.A.S. Abuzaid, and R.M. Benladen, "Opportunities and Challenges of the Knowledge Management Approach to E-Learning: A Case Study in Al-Bayan Model School for Girls, Kingdom of Saudi Arabia," The Electronic Journal on Information Systems in Developing Countries, vol. 35, no. 4, pp. 1-11, 2008.

[13] B. Somekh, "Factors Affecting Teachers' Pedagogical Adoption of Ict," in International Handbook of Information Technology in Primary and Secondary Education, J. Voogt and G. Knezek, Eds., Springer US. p. 449-460, 2008.

[14] R. Jamieson-Proctor, P.C. Burnett, G. Finger, and G. Watson, " Ict Integration and Teachers' Confidence in Using Ict for Teaching and Learning in Queensland State Schools," Australasian Journal of Educational Technology, vol. 22, no. 4, pp. 511-530, 2006.

[15] R. Sani, "Promoting Ict and Learning in Schools :[Main/Lifestyle Edition]", in New Straits Times, April 16. 2007.

[16] B. Cavas, P. Cavas, B. Karaoglan, and T. Kisla, "A Study on Science Teachers' Attitudes toward Information and Communication Technologies in Education," The Turkish Online Journal of Educational Technology, vol. 8, no. 2, pp. 21-31, 2009.

[17] G. Beauchamp, "Teacher Use of the Interactive Whiteboard in Primary Schools: Towards an Effective Transition Framework," Technology, Pedagogy and Education,, vol. 13, no. 3, pp. 327-348, 2004.

[18] E. Neyland, "Integrating Online Learning in Nsw Secondary Schools: Three Schools' Perspectives on Ict Adoption," Australasian Journal of Educational Technology, vol. 27, no. 1, pp. 152-173, 2011.

[19] T.A. Beggs. "Influences and Barriers to the Adoption of Instructional Technology," in Mid-South Instructional Technology Conference. Murfreesboro, TN, 2000.

[20] L.H. Stuart, A.M. Mills, and U. Remus, "School Leaders, Ict Competence and Championing Innovations," Computers & Education, vol. 53, no. 3, pp. 733-741, 2009.

[21] P. Hubber, G. Chittleborough, C. Campbell, W. Jobling, and R. Tytler, "Supporting Ict Based Pedagogies in Science in Rural School Settings " Australian Educational Computing, vol. 25, no. 2, pp. 12-16, 2010.

[22] S.M. Sait, K.M. Al-Tawil, S. Sanaullah, and M. Faheemuddin, "Impact of Internet Usage in Saudi Arabia: A Social Perspective," IJITWE, vol., 2006.

[23] D.N.A. Hayes, "Ict and Learning: Lessons from Australian Classrooms," Computers & Education, vol. 49, pp. 385–395, 2007.

[24] A.G. Almekhlafi, "Preservice and Inservice Teachers, Computer Use in the United Arab Emirates," College of Education, vol. 1, no. 21, pp. 1-34, 2004.

[25] D. Sime and M. Priestley, "Student Teachers' First Reflections on Information and Communications Technology and Classroom Learning: Implications for Initial Teacher Education," Journal of Computer Assisted Learning, vol. 21, no. 2, pp. 130–142, 2005.

[26] A.A. Oyaid, "Education Policy in Saudi Arabia and Its Relation to Secondary School Teachers' Ict Use, Perceptions, and Views of the Future of Ict in Education". 2009, University of Exeter.

[27] P. Giavrimis, S. Giossi, and A. Papastamatis, "Teachers' Attitudes Towards Training in Ict: A Critical Approach," Quality Assurance in Education, vol. 19, no. 3, pp. 283-296, 2011.

[28] V. Cartwright and M. Hammond, "'Fitting It In': A Study Exploring Ict Use in a Uk Primary School," Australasian Journal of Educational Technology, vol. 23, no. 3, pp. 390-407, 2007.

[29] M. Walia, E. Yu, M. Iskander, V. Kapila, and N. Kriftcher, "The Modern Science Lab: Integrating Technology into the Classroom Is the Solution," in Advances in Computer, Information, and Systems Sciences, and Engineering, K. Elleithy, et al., Eds., Springer Netherlands. p. 358-363, 2006.

[30] Noida (2011) Ict for Better Learning Outcomes. Digital Learning.

[31] P. Gannon-Leary, "Implementing an Ict Centre for School and Community," Curriculum Leadership, vol. 7, no. 20, 2009.

[32] J. Sánchez, Á. Salinas, and J. Harris, "Education with Ict in South Korea and Chile," International Journal of Educational Development, vol. 31, no. 2, pp. 126-148, 2011.

[33] J. Sánchez and A. Salinas, "Ict & Learning in Chilean Schools: Lessons Learned," Computers & Education, vol. 51, no. 4, pp. 1621-1633, 2008.

[34] S. Dutta, Charles El-Hage, K. Sabbagh, and P. Tarazi, "Challenges for Information and Communication Technology Development in the Arab World," in Arab World Competitiveness Report 2002-2003 p. 186–208, 2002.

Automatic Scheme for Fused Medical Image Segmentation with Nonsubsampled Contourlet Transform

Ch.Hima Bindu[1]
[1]Associate.Professor, ECE Department, QIS College of Engineering & Technology, Ongole, Andhra Pradesh,

Dr.K.Satya Prasad[2]
[2]Rector & Professor of ECE Department, JNTU Kakinada, Kakinada, Andhra Pradesh, India.

***Abstract*— Medical image segmentation has become an essential technique in clinical and research- oriented applications. Because manual segmentation methods are tedious, and semi-automatic segmentation lacks the flexibility, fully-automatic methods have become the preferred type of medical image segmentation. This work proposes a robust fully automatic segmentation scheme based on the modified contouring technique. The entire scheme consists of three stages. In the first stage, the Nonsubsampled Contourlet Transform (NSCT) of image is computed. This is followed by the fusion of coefficients using fusion method. For that fused image local threshold is computed. This is followed by the second stage in which the initial points are determined by computation of global threshold. Finally, in the third stage, searching procedure is started from each initial point to obtain closed-loop contours. The whole process is fully automatic. This avoids the disadvantages of semi-automatic schemes such as manually selecting the initial contours and points.**

Keywords- Non Sub sampled Contourlet Transform; Image Fusion; Automatic Segmentation.

I. INTRODUCTION

In the recent years, medical image segmentation is the main research subject of image processing applications. This performs various types of volumetric and shape comparisons in the middle of different structures. An application of medical image segmentation with respect to segmenting brain is used in research to characterize neurological diseases such as; multiple sclerosis, schizophrenia and Alzheimer's. Accurate brain segmentation provides volume measurements that can detect the onset of degenerative diseases. Medical imaging is separated into structural and functional systems.

Particularly, applications of medical image analysis, video compression, pattern recognition, etc. are explored by using image segmentation schemes. In general, segmentation schemes could be categorized into two principal types: semi-automatic segmentation [1-2] and fully automatic segmentation [7-8]. Semi-automatic segmentation requires selection of initial points for different images.

The active contour scheme proposed by c.xu et al. is based on the initial contour to obtain the correct contour by minimizing local energy function [9-10]. The Falcao et al. developed the LWOF (Live Wire on the Fly) scheme [11] to select the initial point close to the center. The fully automatic segmentation is applied on body parts such as; leg bones [12], brain [13], fingers [14] or ribs [15] to get their contours. The drawback related to this content of images must be known earlier. The robust fully automatic scheme by Yuan et al. is based on modified edge following technique [16], but the computational time of this scheme is increased. Since this segmentation scheme does not need human input to select initial points and threshold values [16-17]. The conventional edge-following technique only analyzes a given current point and next highest point without considering the neighboring points [18]. In this proposed method the fully automatic segmentation is applied on fused multi source medical images to obtain correct contour points by considering all neighboring pixel values of an image.

Image fusion is the process of combining information from two or more images of a scene into a single composite image which is more informative and is more suitable for visual perception or computer processing. The objective in the image fusion is to reduce irresolution and minimize redundancy in the output, while maximizing relevant information particular to an application or task as well.

Given the same set of input images, different fused images may be created depending on the specific application and what is considered as relevant information. There are several benefits in using image fusion: wider spatial and temporal coverage with decreased irresolution, improved reliability and increased robustness of system performance [2].

So far, several fusion algorithms based on multi source medical images have been proposed. The MRI-CT image fusion using edge preserved technique proposed by Xianghi et al, based on multi scale toggle contrast operator [3]. In this paper proposes a new method using NSCT based fusion process for efficient segmentation. The NSCT is a fully shift-invariant, Multiscale and multi direction expansion that has a fast implementation. The performed computer simulation results showed that the proposed technique was quite efficient. The results are better than the DWT based techniques.

The organization of this paper is as follows, the Section 2 describes the Non Subsampled Contourlet Transform. In Section 3 the generic model of image fusion method. In Section 4 the methodology and the implementation of the proposed process is explained. Section 5 the discussion on the experimental results. In the laconic section the paper is concluded.

II. NON SUBSAMPLED CONTOURLET TRANSFORM

In the foremost contourlet transform down samplers and up samplers are presented in both the laplacian pyramid and the Directional Filter Bank (DFB). Thus, it is not shift-invariant, which causes pseudo-Gibbs phenomena around singularities. NSCT is an improved form of contourlet transform. It is motivated to be employed in some applications, in which redundancy is not a major issue, i.e. image fusion. In contrast with contourlet transform, non-subsampled pyramid structure and non-subsampled directional filter banks are employed in NSCT. The non-subsampled pyramid structure is achieved by using two-channel non subsampled 2-D filter banks. The DFB is achieved by switching off the down samplers/up samplers in each two-channel filter bank in the DFB tree structure and up sampling the filters accordingly. As a result, NSCT is shift-invariant and leads to better frequency selectivity and regularity than contourlet transform. Figure.1 shows the decomposition framework of contourlet transform and NSCT [20].

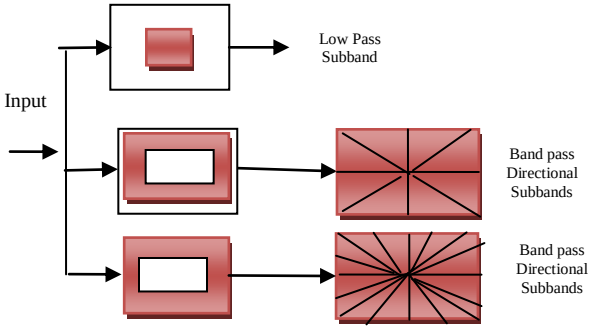


Figure 1: Non Subsampled Contourlet Transform (a) Non Subsampled Filter Bank structure that implements the NSCT.

The NSCT structure consists in a bank of filters that splits the 2-d frequency plane in the subband; these are a non subsampled pyramid structure that ensures the Multiscale property and a non subsampled directional filter bank structure that gives directionality.

III. NSCT BASED FUSION PROCESS

In this paper, image decomposition is performed by the NSCT. The NSCT, which are shift-invariant, multiresolution, localization, directionality, and anisotropy, will be more suitable for image fusion and other image processing, i.e. target recognition, object detection, etc. In the fusion process, both neighborhood coefficients and cousin coefficients information are utilized in the salience measure.

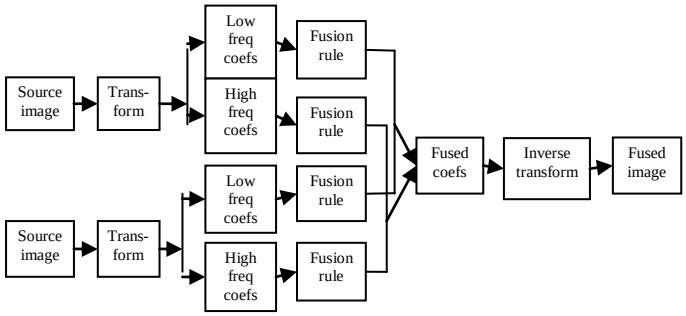


Figure 2: Block diagram of generic model of multi scale image fusion

The Generic model of fusion process is as shown in figure 2 and the rules for fusion of subband coefficients are explained in the following sections:

A. Fusion of Low Frequency subband Coefficients

The coefficients in the coarsest scale sub band represent the approximation component of the source image. It is a smooth and sub sampled version of the original image. Therefore, most of the source images information is kept in low frequency bands. These bands are fused with average selection rule.The proposed selection principles for the sub band coefficients are finally defined as the average selection rule is:

$$I_L^F(i,j)=\begin{cases} I_L^A(i,j) & if: I_L^A(i,j) > I_L^B(i,j) \\ I_L^B(i,j) & if: I_L^A(i,j) \le I_L^B(i,j) \end{cases} \quad (1)$$

B. Fusion of High Frequency subband Coefficients

High-frequency coefficients contain edge and texture features. In order to make full use of information in the neighborhood and cousin coefficients in the NSCT domain, a salience measure, as a spatial frequency of NSCT coefficients is proposed for the first time. Then spatial frequency measures in the overall activity an image is present. Therefore, we propose a scheme by computing the spatial frequency method in a neighbourhood to select the high frequency coefficients.

The spatial frequency (SF), is originated from the HVS, indicates the overall active level in an image and measure the variation of pixels. For an M x N image I, with gray value I (i, j) at position (i, j) the spatial frequency is defined as:

$$SF = \sqrt{(RF)^2 + (CF)^2} \quad (2)$$

Where RF and CF are the row frequency and column frequency respectively:

$$RF = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}\left[I(i,j)-I(i,j-1)\right]^2} \quad 3)$$

$$CF = \sqrt{\frac{1}{MN}\sum_{i=1}^{N}\sum_{j=2}^{M}\left[I(i,j)-I(i-1,j)\right]^2} \quad (4)$$

Each image is partitioned into BxB blocks. The said blocks value varies according to the interest of the user, we consider 8x8 as a block size to obtain more accurate values. Then, compare the spatial frequencies of two corresponding coefficient values in each blocks of I_{HIGH}^A and I_{HIGH}^B to construct the new block of fused image I_{HIGH}^F .

$$I_{HIGH}^F(m,n)=\begin{cases} I_{HIGH}^A(m,n) & if: SF_{HIGH}^A > SF_{HIGH}^B \\ I_{HIGH}^B(m,n) & if: SF_{HIGH}^A < SF_{HIGH}^B \\ \dfrac{I_{HIGH}^A(m,n)+I_{HIGH}^B(m,n)}{2} & otherwise \end{cases} \quad (5)$$

The resultant fused image is compared with basic DWT method. The results are shown below:

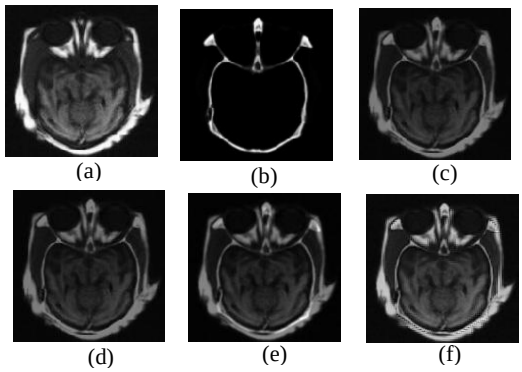


Figure 3 : Mri-Ct Fusion Results :(A) Source Image A (Mri) (B) Source Image (Ct) (C) Ground Truth Image, (D)-(F) Are Output Fused Images By Dwt With Mean Based Method, Dwt With Spatial Frequency Based Method, Nsct With Proposed Method.

IV. PROPOSED METHOD

The entire procedure consists of three stages. The final output is automatic segmented image. The fully automatic segmentation process steps are as follows [19]:

Step1: Read the two different source medical images to be fused (MRI/CT Scan images).

Step2: Perform image fusion on two different images with NSCT fusion process.

Step3: Determine the local threshold

The fused image is partitioned into B X B blocks. The maximum of the difference between the right and left neighboring points in direction of d are defined as $E_{m,n}(i,j)$ in eq. (6).

$$E_{m,n}(i,j) = MAX\left(\left|I^F\left(r^d_{m,n}(i,j)\right) - I^F\left(l^d_{m,n}(i,j)\right)\right|\right) \quad (6)$$

Where d=0 to 3.

Where $r^d_{m,n}(i,j)$ and $l^d_{m,n}(i,j)$ are the right and the left neighboring points of (i, j) in direction of d respectively.

$$T_{gm,n} = \max\left(E_{m,n}(i,j)\right) \quad (7)$$

$$T_g = \frac{\min(T_{gm,n}) + \max(T_{gm,n})}{2} \quad (8)$$

The coordinates of each block in an image frame are (m, n) both m and n range from 0 to B-1. The coordinates in each block are (x, y), where x ranges from 0 to (M/B)-1 and y ranges from 0 to (N/B)-1. M and N represent the width and the height of the image, respectively.

Step 4: Finding the global threshold for entire image using equation (12) after T is computed, Here the m ranges from 0 to M-1 and n ranges from 0 to N-1.This indicates the whole image to compute the relation between all possible neighboring directions is considered. i.e., $E(i,j)$.

Step 5: Obtain the closed loop contours if $E(i,j)$ is greater than T, then the pixel value is set at 1 otherwise 0. This gives segmented output.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The most essential task of the segmentation process is the discrimination of each spot's foreground from its background. The experiment is performed on different MRI & CT scan images of body. The same experiment also performed on individual medical images with sub band coefficient fusion method. The all segmented results give almost closed contours and gives detailed analysis. The various segmented results are shown below which are compared with basic DWT method:

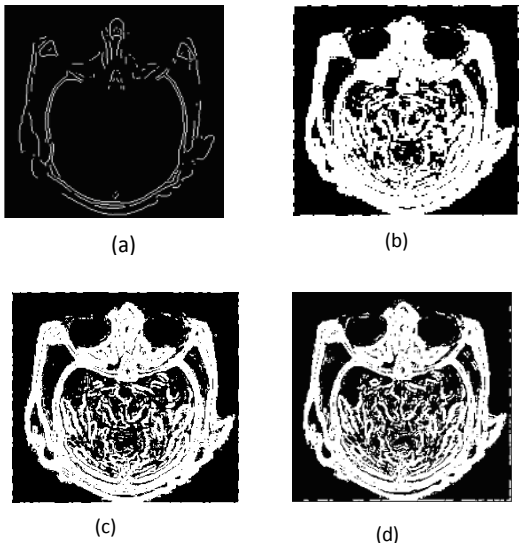


Fig 4: Edge detected outputs of MRI-CT (a) Conventional edge method, (b)-(d) edge detection with Automatic Segmentation approach (b) DWT (fusion with mean values) (c) DWT (fusion with spatial frequency technique) (d) NSCT method (proposed method).

A. Mutual Information (MI)

It is a metric defined as the MI between each source image and fused image. Considering two source images X & Y and fused image Z.

$$I_{Z,X}(Z,X) = \sum P_{Z,X}(Z,X)\log\frac{P_{Z,X}(Z,X)}{P_Z(z)P_X(x)} \quad (9)$$

$$I_{Z,Y}(Z,Y) = \sum P_{Z,Y}(Z,Y)\log\frac{P_{Z,Y}(Z,Y)}{P_Z(z)P_Y(y)} \quad (10)$$

Where P_X, P_Y and P_Z are probability density function in the images X, Y and Z respectively .$P_{Z,X}$ and $P_{Z,Y}$ are joint probability functions. Thus the image fusion performance measure can be defined as

$$MI = I_{Z,X}(Z,X) + I_{Z,Y}(Z,Y) \quad (11)$$

The larger the value of mutual information the better is the fusion result.

B. Peak Signal to Noise Ratio

The more subjective qualitative measurement of distortion is the Peak Signal –to-Noise Ratio (PSNR).It uses a constant

value in which to compare the against instead of a fluctuating signal as in SNR.

$$PSNR = 10\log\left|\frac{255^2}{\frac{1}{MN}\sum\sum(G(m,n)-Z(m,n))^2}\right| \quad (12)$$

TABLE 1: EVALUATION OF FUSED IMAGES

Algorithm	PSNR (db)	MI (bits)
DWT (fusion with mean values)	15.739	1.11
DWT (fusion with Spatial frequency technique)	24.59	2.458
NSCT Method (proposed method)	**25.076**	**2.801**

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities.

For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

CONCLUSION

In this paper, the improved approach of a fully automatic segmented method without any manual interaction was proposed. This method integrates multi scale image segmentation and a statistical fusion scheme. With this proposed method computing and analyzing the characteristics of the left and right neighboring points of the next estimated contour points has the ability to overcome noise interference.

One of the main importance of proposed scheme does not need human input to select initial points and threshold values. This method can easily be extended to locate the closed loop contours of general images with small change on setting the threshold value. This fully automatic segmentation scheme as proposed herein can accurately and repeatedly segment multiple objects for various image and video applications.

REFERENCES

[1] Y. B. Chen, T.-C. Chen, Semi-Automatic image segmentation using dynamic direction prediction, Proc. Of IEEE ICASSP 2002, Vol. 4, May 2002, pp.3369-3372.

[2] K.Karsch, Q.He, Y.Duan, A fast, semi-automatic brain structure algorithm for magnetic resonance imaging, Proc. Of IEEE BIBM 2009, November 2009, pp.297-302.

[3] Sabalan Daneshvar, Hassan Ghassemian, MRI and PET Image Fusion by Combining HIS and retina – inspired Models, Information Fusion 11, 2010, pp.114-123.

[4] Guest Editorial, Image Fusion: Advances in the State Of The Art, Information Fusion 8, 2007, pp.114-118.

[5] Xiangzhi Bai, Fugen Zhou, Bindang Xue, Edge Preserved Image Fusion Based on Multiscale Toggle Contrast Operator, Image And Vision Computing 29 , 2011, pp. 829-839.

[6] Yuhui Liu, Jinzhu Yang, Jinshan Sun, PET/CT Medical Image Fusion Algorithm Based On Multiwavelet Transform, 2nd International conference on advanced computer and control, 2010, pp.264-268.

[7] Y. B. Chen, T.-C. Chen, Semi-Automatic image segmentation using dynamic direction prediction, Proc. Of IEEE ICASSP 2002, Vol. 4, May 2002, pp.3369-3372.

[8] K.Karsch, Q.He, Y.Duan, A fast, semi-automatic brain structure algorithm for magnetic resonance imaging, Proc. Of IEEE BIBM 2009, November 2009, pp.297-302.

[9] Y.B. Chen, O.T.C. Chen, Image segmentation method using thresholds automatically determined from picture contents, EURASIP j.Image Video Process, 2009 (15).

[10] Y.B.Chen,O.T.C.Chen, High accuracy moving object extraction using background subtraction, ICIC Express Lett., December 2009 , pp.649-652.

[11] C.Xu, J.L. Prince, Snakes, Shapes and Gradient Vector Flow, IEEE Trans. Image Process, 7 March 1998, pp.321-331.

[12] W.N.Lie, C.H.Chaung, Fat and accurate snake model for object contour detection, Electron Lett. 37(10), May 2010 (624-626).

[13] A.X.Falco, J.K.Udupa, F.K. Miyazawa, An ultra fast user steered image segmentation paradigm: live wire on the fly, IEEE Trans. Med. Image. 19(1), January 2000, pp.56-62.

[14] Y.B.Chen, Oscal T-C Chen, H.T.Chang, J.T. Chein, An automatic medical assistance diagnosis system applicable on x-ray images. Proc. Of IEEE MWCAS August 2001, pp.910-914.

[15] G.B.A boutanos, J.Nikanne, N.Watkins, B.M. Dawant, Model creation and deformation for the automatic segmentation of the brain in MRI images, IEEE Trans. Biomed. Eng. 46(11),

[16] D.J.Michal, .C. Nelson, HANDX: A model based system for automatic segmentation of bones from digital hard radiog raphs, IEEE Trans. Med. Imag November 1999 (1346-1356) 8(1), March 1989, pp. 64-69.

[17] Z.Yue, A. Goshtasby, L.V.Ackerman, Automatic detection of rib borders in chest radiographs, IEEE Trans. Med. Image 14(3), September 1995, pp.525-536.

[18] Yuan Been Chen, A robust fully automatic scheme for general image segmentation, Digital Signal Processing (21), 2011, pp.87-99.

[19] Ch. Hima Bindu, Dr.K.Satya Prasad, A Fully Automatic Scheme for Medical Image Segmentation With Wavelet Based Image Fusion, Proceedings of the 2011 International Conference on Image Processing, Computer Vision & Pattern Recognition, July 2011, pp.230-235.

[20] Arthur L.da Cuncha,JIANPING Zhou, "The Non subsampled Contourlet transform: Theory, Design and Applications." IEEE Transactions on Image Processing, Vol 15, Oct2006.

AUTHORS PROFILE

Ch.Hima bindu is currently working as Associate Professor in ECE Department, QIS College of Engineering & Technology, ONGOLE, and Andhra Pradesh, India. She is working towards her Ph.D. at JNTUK, Kakinada, India. She received her M.Tech. from the same institute. She has ten years of experience of teaching undergraduate students and post graduate students. She has published 10 research papers in International journals and more than 8 research papers in National & International Conferences. Her research interests are in the areas of image Segmentation, image Feature Extraction and Signal Processing.

Dr.K.Satya Prasad is currently Rector and Professor in ECE Department, JNTUK, Kakinada, India. He received his Ph.D. from IIT, Madras. He has more than 32 years of experience in teaching and 25 years of R & D. He is an expert in Digital Signal Processing. He guided 10 PhD's and guiding 10 PhD scholars. He authored Electronic Devices and Circuits, Network Analysis and Signal & Systems text books. He held different positions in his carrier like Head of the Department, Vice Principal, Principal for JNTU Engg College and Director of Evaluation & presently the Rector of JNTUK.. He published more than 100 technical papers in national and International journals and conferences. The area of interest includes Digital Signal Processing, Image Processing, Communications etc.

Performance Analysis Of Multi Source Fused Medical Images Using Multiresolution Transforms

Ch.Hima Bindu[1]
[1] Assoc. Professor, ECE Department, QIS College of Engineering & Technology, Ongole, Andhra Pradesh,

Dr K.Satya Prasad[2]
[2]Rector& Professor of ECE Department, JNTUK, Kakinada, Andhra Pradesh, India.

***Abstract*— Image fusion combines information from multiple images of the same scene to get a composite image that is more suitable for human visual perception or further image-processing tasks. In this paper the multi source medical images like MRI (Magnetic Resonance Imaging), CT (computed tomography) & PET (positron emission tomography) are fused using different multi scale transforms. We compare various multi resolution transform algorithms, especially the latest developed methods, such as; Non Subsampled Contourlet Transform, Fast Discrete Curvelet, Contourlet, Discrete Wavelet transform and Hybrid Method (combination of DWT & Contourlet) for image fusion. The fusion operations are performed with all Multi resolution transforms. The fusion rules like local maxima and spatial frequency techniques are used for selection in the low frequency and high frequency subband coefficients, which can preserve more information and quality in the fused image. The fused output obtained after the inverse transform of fused sub band coefficients. The experimental results show that the effectiveness of fusion approaches in fusing multi source images.**

Keywords- Image Fusion; Discrete Wavelet Transform; Contourlet Transform; Fast Discrete Curvelet Transform; Nonsubsampled Contourlet Transform.

I. INTRODUCTION

The Image fusion is the process of combining information from two or more images of a scene into a single composite image which is more informative and is more suitable for visual perception or computer processing. The objective in the image fusion is to reduce uncertainty and minimize redundancy in the output, while maximizing relevant information particular to an application or task as well. Given the same set of input images, different fused images may be created depending on the specific application and what is considered as relevant information. There are several benefits in using image fusion: wider spatial and temporal coverage with decreased uncertainty, improved reliability and increased robustness of system performance [Ref.3].

The medical images like MRI and CT provides high-resolution images with structural and anatomical information. PET images provide functional information with low spatial resolution. In the recent years, the success of MRI-CT, PET-MRI [Ref.1] & PET-CT [Ref.6] imaging in the clinical field triggered considerable interest in noninvasive functional and anatomical imaging. The limited spatial resolution in PET images is often resulted unsatisfactory in morphological analysis. Combining anatomical and functional tomography datasets provide much more qualitative detection and quantitative determination in this area [Ref .1, Ref .10].

So far, several fusion algorithms based on multi source medical images have been proposed. The MRI-CT image fusion using edge preserved technique proposed by Xianghi et al, based on multi scale toggle contrast operator [Ref .4]. The MRI-PET image fusion proposed by Sabalan Daneshvar et al based on combining HIS (Hue Intensity Saturation model) and retina models improve the functional and spatial information content [Ref .1]. The CT-PET image fusion proposed by Yuhui Liu et al, based on multi wavelet transform adds more details and structure information [Ref .6].

The rest of the paper is organized as follows: Section 2 explains Multiscale transforms. Section 3 presents generic fusion model. Section 4 explains the proposed method. Section 5 the discussion on the experimental results. In the laconic section, the paper is concluded.

II. MULTIRESOLUTION TRANSFORMATIONS

A. Discrete Wavelet Transform

Discrete Wavelet transform (DWT) provides a framework in which a signal is decomposed, with each level corresponding to lower frequency sub band, and higher frequency sub bands. There are two main groups of transforms: continuous and discrete. In one dimension the idea of the wavelet transform is to present the signal as a superposition of wavelets. If a signal is represented by f (t), the wavelet decomposition is

$$f(t) = \sum_{m,n} c_{m,n} \psi_{m,n}(t) \tag{2.1}$$

Where $\psi_{m,n}(t) = 2^{-m/2} \psi\left(2^{-m} t - n\right)$, m and n are integers. There exist very special choices of ψ such that $\psi_{m,n}(t)$ constitutes an ortho normal basis, so that the wavelet transform coefficient can be obtained by an inner calculation:

$$c_{m,n} = \langle f, \psi_{m,n} \rangle = \int \psi_{m,n}(t) f(t) dt \tag{2.2}$$

In order to develop a multiresolution analysis, a scaling function ϕ is needed, together with the dilated and translated parameters of $\phi_{m,n}(t) = 2^{-m/2} \phi\left(2^{-m} t - n\right)$. The signal $f(t)$ can be decomposed in its coarse part and details of various sizes by projecting it onto the corresponding spaces. Therefore, the approximation coefficients $a_{m,n}$ of the

function f at resolution 2^m and wavelet coefficients $c_{m,n}$ can be obtained:

$$a_{m,n} = \sum_k h_{2n-k} a_{m-1,k} \quad (2.3)$$

$$C_{m,n} = \sum_k g_{2n-k} a_{m-1,k} \quad (2.4)$$

Where h_n is a low pass FIR filter and g_n is a high pass FIR filter. To reconstruct the original signal, the analysis filter can be selected from a biorthogonal set which have a related set of synthesis filters. These synthesis filters $\tilde{h}$ and $\tilde{g}$ can be used to perfectly reconstruct the signal using the reconstruction formula

$$a_{m-1,l}(f) = \sum_n \left[\tilde{h}_{2n-l} a_{m,n}(f) + \tilde{g}_{2n-l} c_{m,n}(f) \right] \quad (2.5)$$

Equations (2.3) and (2.4) are implemented by filtering and down sampling. Conversely equation (2.5) is implemented by an initial up sampling and a subsequent filtering.

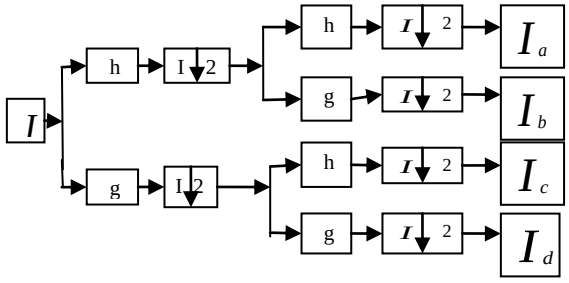


Fig 1: Structure of 2-D DWT

In a 2-D DWT, a 1-D DWT is first performed on the rows and then columns of the data by separately filtering and down sampling. This result in one set of approximation coefficients I_a and three set of detail coefficients, as shown in Fig 1, where I_b, I_c, I_d represent the horizontal, vertical and diagonal directions of the image I, respectively.

In the filter theory, these four sub images correspond to the outputs of low-low (**LL**), low-high (**LH**), high-low (**HL**), and high-high (**HH**) bands. By recursively applying the same scheme to the LL sub band multi resolution decomposition with a desire level can then be achieved. There, a DWT with K decomposition levels will have M=3*K+1 such frequency bands.

B. *Contourlet Transform*

The contourlet transform is first developed in continuous domain and then is discretized for sampled data; contoured transform starts with a discrete domain construction. The contourlet is also deemed as a "true" two dimensional transform that can capture the intrinsic geometrical structure of an image. Two filter banks are employed to implement the contoured transform as shown in Fig.2.

The Laplacian pyramid is first used to capture the point discontinuities, and then a directional filter bank is used to link point discontinuities into linear structures. As the DWT, the contoured transform also has no shift invariant property because of the down-sampling operation.

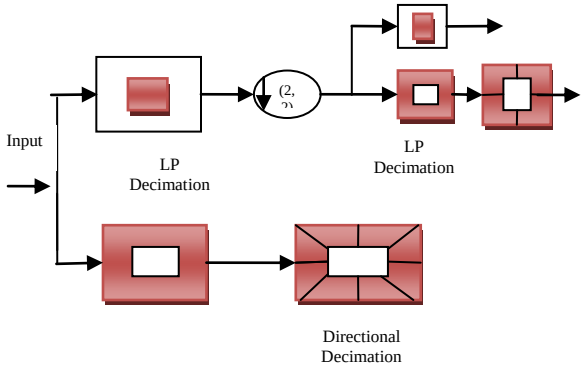


Fig 2: Block diagram of contourlet transform

Do and Vetterli found that to obtain a sparser representation for 2-D piecewise smooth functions in R2, an effective method is to utilize a double filter bank scheme, that is, first apply a multiscale decomposition to capture point discontinuities and then perform a local directional decomposition to synthesize the nearby edge points into independent contour segments.

With a rich set of basis oriented at various directions and scales, contourlet can effectively capture the intrinsic contours and edges in natural images that set radiational multiresolution analysis methods are difficult to handle. Contourlet offer s a much richer sub band set of different directions and shapes, which helps to capture geometric structures in images much more efficiently [Ref.7].

C. *Non Subsampled Contourlet Transform (NSCT)*

In the foremost contourlet transform down samplers and up samplers are presented in both the laplacian pyramid and the Directional Filter Bank (DFB). Thus, it is not shift-invariant, which causes pseudo-Gibbs phenomena around singularities. NSCT is an improved form of contourlet transform. It is motivated to be employed in some applications, in which redundancy is not a major issue, i.e. image fusion. In contrast with contourlet transform, non subsampled pyramid structure and non subsampled directional filter banks are employed in NSCT. The non subsampled pyramid structure is achieved by using two-channel non subsampled 2-D filter banks. The DFB is achieved by switching off the downsamplers/up samplers in each two-channel filter bank in the DFB tree structure and up sampling the filters accordingly. As a result, NSCT is shift-invariant and leads to better frequency selectivity and regularity than contourlet transform. Fig.3 shows the decomposition framework of contourlet transform and NSCT [Ref.11]. The NSCT structure consists in a bank of filters that splits the 2-D frequency plane in the subband; these are a non subsampled pyramid structure that ensures the Multiscale property and a non subsampled directional filter bank structure that gives directionality [Ref.15].

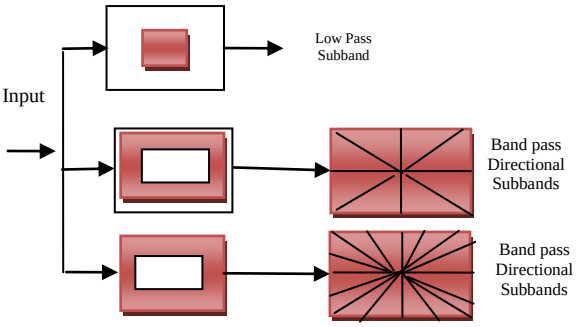


Fig 3: Block diagram of NSCT

A. Curvelet Transform

The DWT, SWT, and DTCWT cannot capture curves and edges of images. More reasonable bases should contain geometrical structure information when they are used to represent images. Candes and Donoho proposed the Curvelet transform (CVT) with the idea of representing a curve as a superposition of bases of various lengths of width obeying the scaling width ≈ length . The CVT is referred to as the ‘‘true” 2D transform. The discrete version implemented in this is a ‘‘wrapping” transform. The second generation of Curvelet transform is presented in Fig: 4. Firstly, the 2D FFT is applied to the source image to obtain Fourier samples. Next, a discrete localizing window smoothly localizes the Fourier transform near the sheared wedges obeying the parabolic scaling. Then, the wrapping transformation is applied to re-index the data. Finally, the inverse 2D FFT is used to obtain the discrete CVT coefficients [Ref.16].

Fig 4: The second generation of Curvelet transforms

III. HYBRID METHOD

The hybrid method can be referred from [Ref.2, Ref.17] references. In this paper reference [Ref.17] method is considered as basic method to do further comparisons.

IV. GENERIC MODEL OF MULTISCALE-BASED IMAGE FUSION

In this paper, there are two different input medical source images A and B (like MRI & CT, MRI & PET, CT & PET). The image fusion algorithm should preserve all the salient features of source images. Fig 5 illustrates the generic image fusion frame work based on Multiscale image decomposition methods. The source images are firstly decomposed into low-frequency sub bands and a sequence of high-frequency sub bands in different scales and orientations. Then the fusion coefficients are obtained from sub bands according to fusion rules. Finally, fused image is reconstructed by applying inverse transform on the fused sub bands.

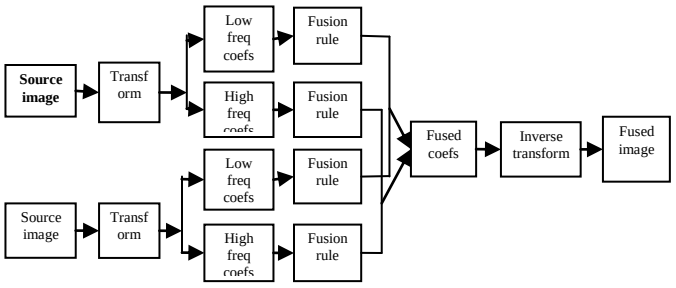


Fig 5: Block diagram of generic model of multi scale image fusion

The key issue in spatial domain algorithms is identifying the most important information in source images and fusing the salient information into the fused image [Ref.8]. This paper includes the average pixel method and PCA method. The fusion rules are explained in the following sections.

A. Fusion rule of lower sub band coefficients

The coefficients in the coarsest scale sub band represent the approximation component of the source image. It is a smooth and sub sampled version of the original image. Therefore, most of the source images information is kept in low frequency bands. The proposed selection principles for the sub band coefficients are finally defined as the maximum selection rule is:

$$I_L^F(i,j)=\begin{cases} I_L^A(i,j) & if: I_L^A(i,j) > I_L^B(i,j) \\ I_L^B(i,j) & if: I_L^A(i,j) \le I_L^B(i,j) \end{cases} \quad (4.1)$$

B. Fusion rule of higher sub band coefficients

The highpass sub band coefficients represent the detailed components of the source image; according to characteristic of HVS. So, it is easy to find that for the high resolution region the human visual interest is concentrated on the detection of changes in between regional contrast. Then spatial frequency measures in the overall activity an image is present. Therefore, we propose a scheme by computing the spatial frequency method in a neighbourhood to select the high frequency coefficients. The spatial frequency (SF), is originated from the HVS, indicates the overall active level in an image and measure the variation of pixels [9]. For an M x N image I, with gray value I (i, j) at position (i, j) the spatial frequency is defined as

$$SF=\sqrt{(RF)^2+(CF)^2} \quad (4.2)$$

Where RF and CF are the row frequency and column frequency respectively:

$$RF=\sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=2}^{N}\left[I(i,j)-I(i,j-1)\right]^2} \quad (4.3)$$

$$CF=\sqrt{\frac{1}{MN}\sum_{i=1}^{N}\sum_{j=2}^{M}\left[I(i,j)-I(i-1,j)\right]^2} \quad (4.4)$$

Each image is partitioned into BxB blocks. The said blocks value varies according to the interest of the user, we consider 8x8 as a block size to obtain more accurate values. Then, compare the spatial frequencies of two corresponding coefficient values in each blocks of I^{A}_{HIGH} and I^{B}_{HIGH} to construct the new block of fused image I^{F}_{HIGH} .

$$I^{F}_{HIGH}(m,n)=\begin{cases} I^{A}_{HIGH}(m,n) & if: SF^{A}_{HIGH} > SF^{B}_{HIGH} \\ I^{B}_{HIGH}(m,n) & if: SF^{A}_{HIGH} < SF^{B}_{HIGH} \\ \dfrac{I^{A}_{HIGH}(m,n)+I^{B}_{HIGH}(m,n)}{2} & otherwise \end{cases} \quad (4.5)$$

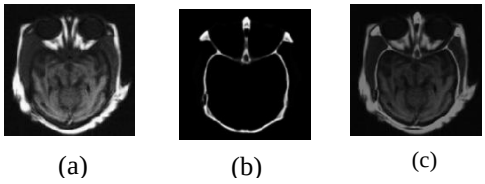
V. PROPOSED METHOD

The performance comparison of Multiscale transform method has accomplished the following Steps:

- Decompose the source images A and B into low frequency subband and a series of high frequency sub bands at L levels by using various Multiscale transforms.
- Fuse the low frequency subband coefficients and high frequency subband coefficients according to lower subband fusion rule equation (4.1) and higher subband fusion rules equation (4.2-4.5).
- Reconstruct the original image based on the new fused coefficients of sub bands by taking respective inverse Multiscale Transform, thus fused image is obtained.
- Compare the obtained fused image from various transforms with different evaluation parameters from equation (6.1-6.10).

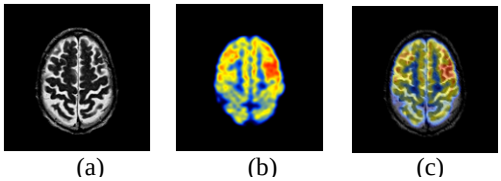
VI. EXPERIMENTAL RESULTS AND ANALYSIS

To evaluate the performance of the multi scale transforms, several experimental results are presented in this section. Experiments are performed on three different multisource images MRI, CT & PET. The MRI and PET images are downloaded from the Harvard university site (http:// www.med.harvard.edu/ AANLIB/ home.html). Similarly the CT and MRI images are downloaded from www.imagefusion.org website link. The proposed method is applied to these image sets. To show the effectiveness of the multi scale transform the comparisons start with simple basic average method, PCA (principal component analysis) method and Brovery method under spatial domain techniques and DWT [Ref.13-14], Contourlet Transform[Ref.5], Hybrid Method is combination of DWT & Contourlet [Ref.2, Ref.12] and Non subsampled contourlet transform in transform domain techniques. The fused image output based on different methods is shown from Fig 6 – 8. The performances of fused medical images are shown in Fig: 9 using Multiscale transform with different evaluation parameters.



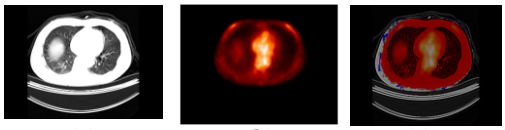
(a) (b) (c)

Fig 6: MRI-CT fusion results (a) Source image A (MRI) (b) Source image (CT) (c) Fused image.



(a) (b) (c)

Fig 7: MRI-PET fusion results (a) Source image A (MRI) (b) Source image B (PET) (c) Fused image.



(a) (b) (c)

Fig 8: CT-PET fusion results: (a) Source image A (CT) (b) Source image (PET) (c) Fused image.

A. Peak Signal to Noise Ratio (PSNR)

PSNR (Peak Signal to Noise Ratio) is a metric for the ratio between the maximum possible power of a signal and power of corrupting noise that affects the fidelity of its representation. It is used to measure the quality of reconstructed images.

$$PSNR = 10\log_{10}\left(\frac{L^2}{\frac{i}{M\times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(R(i,j)-I^{F}(i,j)\right)^2}\right) \quad (6.1)$$

Where $R(i,j)$ and $I^{F}(i,j)$ are the pixel values of the ideal reference and the obtained fused image, respectively. M and N are the dimensions of the images.

B. Mutual Information (MI)

Mutual information of two random variables is a quantity that measures the mutual dependence of the two variables. Here, $MI_{RI}{}^{F}$ measures the information that reference and the fused image shares:

$$MI_{RIF} = \sum_{i=1}^{L}\sum_{j=1}^{L} P_{RIF}(i,j)\log_2\frac{P_{RIF}(i,j)}{P_R(i)P_{IF}(j)} \quad (6.2)$$

Where P_{RI} is the normalized joint gray level histogram of images R and I^{F}, P_R and P_{I^F} are the normalized marginal histograms of the two images.

The mutual information I_{AF} between the sources images A and the fused image F is defined as follows:

$$I_{AF} = \sum_{AF} P_{AF}(a, f) \log \frac{P_{AF}(a, f)}{P_A(a) P_F(f)} \quad (6.3)$$

Where P_{AF} is the jointly normalized histogram of A and F, P_A and P_F are the normalized histogram of A and F, and a f represent the pixel value of the image A and F, respectively. The mutual information I_{BF} between the source image B and the fused image F are similar to I_{AF} . The mutual information between the source images A, B, and the fused image F is the sum of I_{AF} and I_{BF} , i.e.

$$MI_F^{AB} = I_{AF} + I_{BF} \quad (6.4)$$

C. Gradient and Wrap

The gradient value and wrap values are facilitate the correlation between the resultant F and reference R images. If gradient value is high and wrap value is low then the two images are more correlated. The gradient and warp are defined as follows:

$$Grad = 1/M * N \frac{\sum\sum \sqrt{[F(i,j) - F(i+1,j)]^2 + [F(i,j) - F(i,j+1)]^2}}{\sqrt{2}} \quad (6.5)$$

$$Wrap = 1/M * N \sum_{i=1}^{M} \sum_{j=1}^{N} F(i,j) - R(i,j) \quad (6.6)$$

Where M and N are size of the image.

D. Effective Great Degrees

This is because that the metrics $Q^{AB/F}$ and Q_0 mainly measure the amount of salient information transferred from source images into the fused image [Ref.17]. The metric $Q^{AB/F}$ is defined as follows:

$$Q^{AB/F} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M}\left(Q^{AF}(n,m)W^A(n,m) + Q^{BF}(n,m)W^B(n,m)\right)}{\sum_{n=1}^{N}\sum_{m=1}^{M}\left(W^A(n,m) + W^B(n,m)\right)} \quad (6.7)$$

The dynamic range of $Q^{AB/F}$ is [0 1], and it should be as close to 1 as possible.

Another metric Q_O is as follows:

$$Q_O(A,F) = \frac{2\sigma_{af}\, 2\bar{a}\,\bar{f}}{(\sigma_a^2 + \sigma_f^2)(\bar{a}^2 + \bar{f}^2)} \quad (6.8)$$

Where σ_{af} represents the coherence between A and F, σ_a σ_f denote the standard deviation of A and F: $\bar{a}$, $\bar{f}$ represent the mean value of A and F respectively, similarly calculate Q_o(B,F) using above equation then finally the Q_0 value is as follows:

$$Q_0 = \frac{Q_0(A,F) + Q_0(B,F)}{2} \quad (6.9)$$

The Q_0 range is $-1 \leq Q_0 \leq 1$ and it should be also as close to 1 as possible.

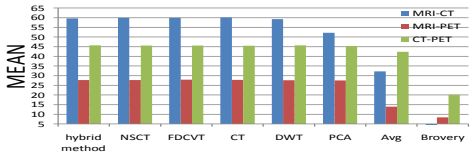
E. Correlation

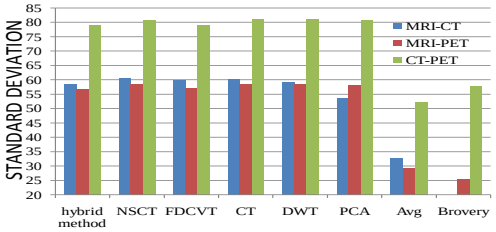
It computes the correlation coefficient between fused image F and reference image R. to computes the correlation coefficient using following equation:

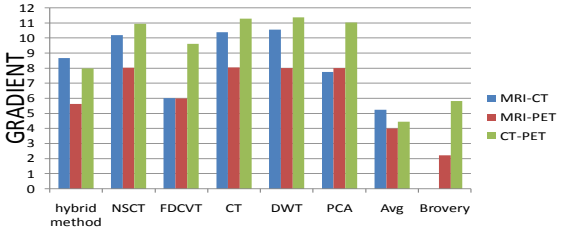
$$corr = \frac{\sum_{i}^{M}\sum_{j}^{N}(F(i,j) - \bar{F})(R(i,j) - \bar{R})}{\sqrt{(\sum_{i}^{M}\sum_{j}^{N}\left(F(i,j) - \bar{F}\right)^2)\sum_{i}^{M}\sum_{j}^{N}(\left(R(i,j) - \bar{R}\right)^2)}} \quad (6.10)$$

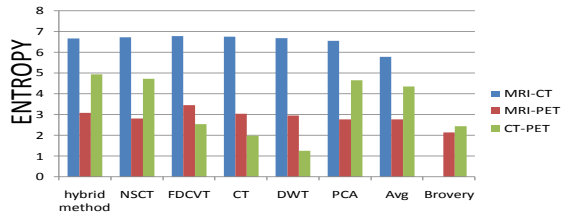
Where $\bar{F}$ and $\bar{R}$ are mean values of fused and reference images respectively.

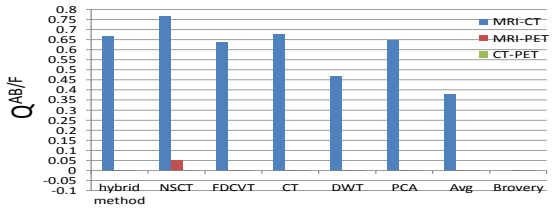
For Table I-III refer Appendix A and The graphical representation of all performance evaluation parameters are hown below (Fig 9):

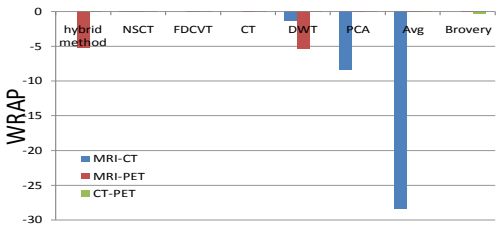


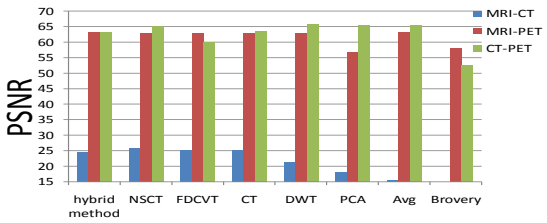


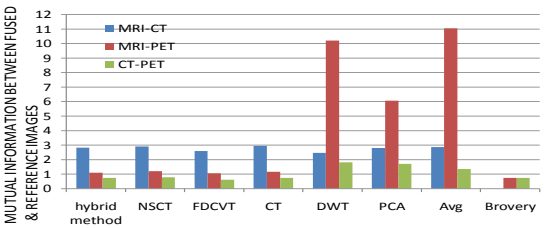












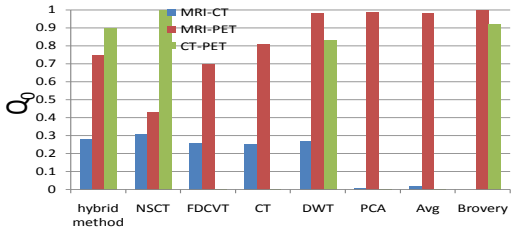
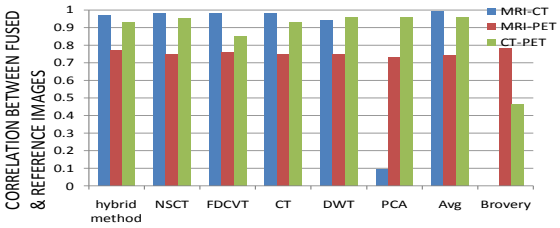
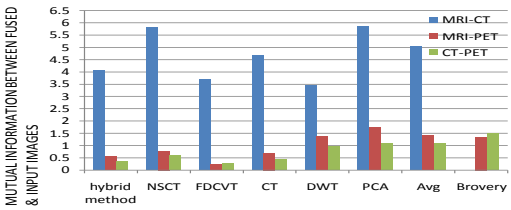


Fig: 9: Performance of fused medical image using different Multiscale transforms with different evaluation parameters.





VII. Conclusion

In this paper, performance analysis is compared with different multi scale transforms for the fusion of multi source images. The fusion procedure is discussed with different subband fusion rules to obtain fused image from different multi sources. The obtained fusion results are compared with different evaluation parameters. The comparison of transform methods are starts with those of the pixel averaging, PCA method and Brovery methods. Finally those compared with DWT, Contourlet, Hybrid Method (is combination of DWT & Contourlet), Fast Discrete Curvelet and NSCT fusion methods. We calculated and analyzed tools such as; PSNR, MI, Effective great degrees, gradient, wrap, standard deviation, mean and entropy values.

References

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. (references) Sabalan Daneshvar, Hassan Ghassemian, MRI and PET Image Fusion by Combining HIS and retina – inspired Models, Information Fusion 11, 2010, pp.114-123.

[2] Shutao Li, Bin Yang, Multifocus Image Fusion by Combining Curvelet and Wavelet Transform, Pattern Recognition Letters 29 , 2008, pp.1295-1301.

[3] Guest Editorial, Image Fusion: Advances in the State Of The Art, Information Fusion 8, 2007, pp.114-118.

[4] Xiangzhi Bai, Fugen Zhou, Bindang Xue, Edge Preserved Image Fusion Based on Multiscale Toggle Contrast Operator, Image And Vision Computing 29, 2011, pp. 829-839.

[5] L.Yang, B.L.Guo, W.Ni, Multimodality Medical Image Fusion based on Multiscale Geometric Analysis of Contourlet Transform, Neurocomputing 72, 2008, pp. 203-211.

[6] G.G.Bhutada, R.S.Anand, S.C. Saxena, Edge Preserved Image Enhancement using Adaptive Fusion of Images Denosied By Wavelet And Curvelet Transform, Digital Signal Processing 21 (2011) 118-130.

[7] Yi Chai, Huafeng Li, Xiaoyang Zhang, Multifocus Image Fusion based on features Contrast of Multiscale products in nonsubsampled Contourlet Transform Domain, Optik 123 2012, pp.569-581.

[8] Xiaoquing Zhang, Yongguo Zheng, Yanjun Peng, Weike Liu, Changqiang Yang, Research on multi-mode medical image fusion algorithm based on wavelet transform and edge characteristics of images, 2nd International conference on Image and signal processing, 2009, pp.1-4.

[9] Yuhui Liu, Jinzhu Yang, Jinshan Sun, PET/CT Medical Image Fusion Algorithm Based On Multiwavelet Transform, 2nd International conference on advanced computer and control, 2010, pp.264-268.

[10] I.Pitas, Digital Image processing Scheme and Application, john wiley & sons. New York, 2000.

[11] A.Soma Sekhar, Dr.M.N. Giri Prasad, A Novel Approach Of Image Fusion On MR And CT Images Using Wavelet Transforms, Proc. Of IEEE ICECIT 3rd International Conference, July 2011, pp. 172-176.

[12] M.N.Do, M.Verrerli, The contourlet transform: an efficient directional multi resolution image representation, IEEE Trans. On Image Processing 14(12), 2005, pp.2091-2106.

[13] G.Piella, A General Framework for Multi resolution image Fusion: From Pixels to Regions. Information Fusion (4), 2003, pp.259-280.

[14] S.T.Li,b.Yang, Multifocus Image Fusion Using Region Segmentation and Spatial Frequency, Image and Vision Computing 26 (7) ,2008, pp. 971-979.

[15] Shutao Li, Bin Yang, Jianwen Hu, Performance comparison of Multi resolution transform for image fusion, Information Fusion 12, 2011, pp.74-84.

[16] Arthur L.da Cuncha,JIANPING Zhou, "The Nonsubsampled Contourlet transform: Theory, Design and Applications." IEEE Transcations on Image Processing, Vol 15, Oct2006.

[17] Ch.Hima Bindu, Dr.K.Satya Prasad, "MRI-PET medical image fusion by combining DWT and contourlet transform." To be published in Springer conference, Aug2012, ITC 2012, LNEE, pp. 124-129, 2012.

[18] Ch.Hima Bindu, et.al, "Discrete Wavelet Transform Based Medical Image Fusion using Spatial Frequency Techniques", International Conference on Recent Advances in Engineering and Technology (ICRAET 2012), Apr 2012.

[19] Ch.Hima Bindu, et.al, " Multimodal Medical Image Fusion of MRI-PET using wavelet transform." To be published in 2012 International Conference on Advances in Mobile Network, Communication and Its Applications Aug2012.

[20] Arthur L.da Cuncha,JIANPING Zhou, "The Non subsampled Contourlet transform: Theory, Design and Applications." IEEE Transactions on Image Processing, Vol 15, Oct2006.

[21] Candès, L. Demanet, D. Donoho, L.X. Ying, Fast discrete curvelet transforms, SIAM Multiscale Modeling and Simulation 5 (3) (2006) pp: 861–899.

[22] C.S. Xydeas, V. Petrovic, Objective image fusion performance measure, Electronic Letters 36 (4) (2000) 308–309.

AUTHORS PROFILE

Ch.Hima bindu is currently working as Associate Professor in ECE Department, QIS College of Engineering & Technology, ONGOLE, and Andhra Pradesh, India. She is working towards her Ph.D. at JNTUK, Kakinada, India. She received her M.Tech. from the same institute. She has ten years of experience of teaching undergraduate students and post graduate students. She has published 10 research papers in International journals and more than 8 research papers in National & International Conferences. Her research interests are in the areas of image Segmentation, image Feature Extraction and Signal Processing.

Dr.K.Satya Prasad is currently Rector and Professor in ECE Department, JNTUK, Kakinada, India. He received his Ph.D. from IIT, Madras. He has more than 32 years of experience in teaching and 25 years of R & D. He is an expert in Digital Signal Processing. He guided 10 PhD's and guiding 10 PhD scholars. He authored Electronic Devices and Circuits, Network Analysis and Signal & Systems text books. He held different positions in his carrier like Head of the Department, Vice Principal, Principal for JNTU Engg College and Director of Evaluation & presently the Rector of JNTUK.. He published more than 100 technical papers in national and International journals and conferences. The area of interest includes Digital Signal Processing, Image Processing, Communications etc.

Appendix A : Performance Analysis Tables

Table: I Evaluation of different methods for CT-MRI images

Algorithm	Mean	Std	Entropy (bits/sec)	Grad	Wrap	MI_{RI^F}	Corr	PSNR (db)	MI_F^{AB} (mean/std)	$Q^{AB/F}$ (mean/std)	Q_0 (mean/std)
Hybrid Approach	59.67	58.60	6.67	8.68	0	2.83	0.97	24.63	4.06	0.67	0.28
Non Subsampled Contourlet Transform	59.96	**60.51**	6.73	10.20	0.98	2.91	0.98	**25.95**	5.82	**0.77**	**0.31**
Fast Discrete Curvelet Transform	59.93	59.81	**6.79**	6.02	**0**	2.58	0.98	25.26	3.70	0.64	0.26
Contourlet	**60.13**	60.44	6.76	10.4	0	**2.94**	0.98	25.23	4.65	0.68	0.25
Discrete Wavele Transform	59.26	59.42	6.68	**10.56**	-1.38	2.46	0.94	21.19	3.43	0.47	0.27
PCA Method	52.16	53.53	6.55	7.75	-8.47	2.79	0.09	17.96	**5.86**	0.65	0.01
Pixel average	32.17	32.60	5.78	5.25	**-28.47**	2.86	**0.99**	15.55	5.05	0.38	0.02

Table: II Evaluation of different methods for MRI-PET images

Algorithm	Mean	Std	Entropy (bits/sec)	Grad	Wrap	MI_{RI}	Corr	PSNR (db)	MI_F^{AB} (mean/std)	$Q^{AB/F}$ (mean/std)	Q_0 (mean/std)
Hybrid Approach	27.70	56.78	3.09	5.64	-5.19	1.10	0.77	**63.15**	0.55	0.001	0.75
Non Subsampled Contourlet Transform	27.69	58.48	2.82	8.03	0.12	1.2	0.75	62.91	0.78	**0.054**	0.43
Fast Discrete Curvelet Transform	**27.90**	56.99	**3.46**	6.00	0.01	1.06	0.76	63.06	0.23	0.001	0.70
Contourlet	27.77	58.43	3.05	**8.06**	0.01	1.15	0.75	62.91	0.7	0.001	0.81
Discrete Wavele Transform	27.65	**58.49**	2.96	8.02	-5.4	10.22	0.75	62.91	1.38	0.001	0.98
PCA Method	27.54	58.27	2.77	8.01	0.18	6.06	0.73	56.72	**1.72**	0.003	0.99
Pixel average	13.92	29.36	2.77	4.01	**-0.05**	**11.06**	0.74	**63.15**	1.42	0.001	0.98
Brovary Method	8.42	25.47	2.15	2.22	0.13	0.74	**0.78**	58.13	1.35	0.001	**1**

Table: III Evaluation of different methods for CT-PET images

Algorithm	Mean	Std	Entropy (bits/sec)	Grad	Wrap	$MI_{RI}{}^{F}$	Corr	PSNR (db)	MI_{F}^{AB} (mean/std)	$Q^{AB/F}$ (mean/std)	Q_0 (mean/std)
Hybrid Approach	**45.65**	79.06	**4.94**	7.99	0.023	0.73	0.93	63.27	0.37	0.0004	0.9
Non Subsampled Contourlet Transform	45.6	80.91	4.73	10.95	0.56	0.77	0.95	65.26	0.59	0.0004	**1**
Fast Discrete Curvelet Transform	45.6	79.17	2.54	9.63	0.06	0.61	0.85	60.01	0.25	0.0005	0.005
Contourlet	45.53	81.05	1.99	11.3	0.029	0.74	0.93	63.51	0.42	0.0003	0.004
Discrete Wavelet Transform	**45.65**	**81.29**	1.25	**11.37**	0.001	**1.81**	**0.96**	**65.7**	0.97	0.0006	0.83
PCA Method	45.29	80.82	4.66	11.04	0.005	1.71	**0.96**	65.37	1.09	0.0006	0.004
Pixel average	42.37	52.09	4.36	4.46	0.007	1.34	**0.96**	65.4	1.09	0.0006	0.004
Brovary Method	20.09	57.68	2.45	5.83	**-0.371**	0.73	0.46	52.44	**1.51**	**0.0014**	0.92

Defending Polymorphic Worms in Computer Network using Honeypot

R. T. Goswami[a], Avijit Mondal[b]
[a,b] Department of Computer Science, Birla Institute of Technology Extension Centre, Kolkata, India–700107

Bimal Kumar mishra[c1], N.C. Mahanti[d]
[a,b] Department of Applied Mathematics, Birla Institute of Technology, Mesra, Ranchi, Indian-835

***Abstract*— Polymorphic worms are a major threat to internet infrastructure security. In this mechanism we are using gate-translator, double honeypot, sticky honeypot, internal translator and antivirus of Cloud AV,which attracts polymorphic worms. We are proposing an algorithm to detect and remove polymorphic worms and innocuous traffic related packets.**

Keywords- Polymorphic worm; Honeypot; Honeynet; Sticky honeypot; Cloud computing.

I. INTRODUCTION

Worms are computer programs that self replicate without requiring any human intervention, by sending copies of their code in network packets and ensuring the code is executed by the computers that receive it. When computers are infected, they spread copies of themselves and perform other malicious activities. A polymorphic worm is a worm that changes its appearance with every instance [1]. There are two basic types of intrusion detection: host-based and network-based. Host-based IDSs examine data held on individual computers that serve as hosts, while network-based IDSs examine data exchanged between computers [3, 4].

Security experts manually generate the IDS signatures by studying the network traces after a new worm has been released. Our research is based on Honeypot technique. Developed in recent years, honeypot is a monitored system on the Internet serving the purpose of attracting and trapping attackers who attempt to penetrate the protected servers on a network. Honeypots fall into two categories. A high-interaction honeypot such as (Honeynet) operates a real operating system and one or multiple applications. A low-interaction honeypot such as (Honyed) simulates one or multiple real systems. In general, any network activities observed at honeypots are considered suspicious [1, 2].

Security experts need a great deal of information to perform signature generation. Such information can be captured by tools such as honeynet. Honeynet is a network of standard production systems that are built together and are put behind some type of access control device (such as a firewall) to watch what happens to the traffic [1]. We assume the traffic captured by honeynet is suspicious. Our system reduces the rate of false alarms by using honeynet to capture traffic destined to a certain network.

The attackers will try every possible way to extend the life time of Internet worms. In order to evade the signature-based system, a polymorphic worm appears differently each time it replicates itself. This subsection discusses the polymorphism of Internet worms. There are many ways to make polymorphic worms [2]. One technique relies on self encryption with a variable key. It encrypts the body of a worm that erases both signatures and statistical characteristics of the worm byte string. A copy of the worm, the decryption routine, and the key are sent to a victim machine, where the encrypted text is turned into a regular worm program by the decryption routine. The program is then executed to infect other victims and possibly damage the local system. If the same decryption routine is always used, the byte sequence in the decryption routine can serve as the worm signature. A more sophisticated method of polymorphism is to change the decryption routine each time a copy of the worm is sent to another victim host. This can be achieved by keeping several decryption routines in a worm. When the worm tries to make a copy, one routine is randomly selected and other routines are encrypted together with the worm body.

The number of different decryption routines is limited by the total length of the worm. Given a limited number of decryption routines, it is possible to identify all of them as attack signatures after enough samples of the worm have been obtained. Another polymorphism technique is called garbage-code insertion. It inserts garbage instructions into the copies of a worm. For example, a number of nop (i.e., no operation) instructions can be inserted into different places of the worm body, thus making it more difficult to compare the byte sequences of two instances of the same worm. However, from the statistics point of view, the frequencies of the garbage instructions in a worm can differ greatly from those in normal traffic. If that is the case, anomaly-detection systems can be used to detect the worm. Furthermore, some garbage instructions such as nop can be easily identified and removed.

A Cloud AV: N-version antivirus identifies malicious software by multiple, heterogeneous engine in parallel to provide N-version protection. Cloud AV includes a light weight, cross platform host agent, with ten antivirus engine and two behavioral detection engines [5].

The attacker sends one instance of a polymorphic worm to a network, and this worm in every infection automatically attempts to change its payload to generate other instances. So, if we need to capture all polymorphic worm instances, we need to give a polymorphic worm chance to interact with hosts without affecting their performance. So, we propose new detection method “Double-honeynet” to interact with polymorphic worms and collect all their instances. The

proposed method makes it possible to capture all worm instances and then forward these instances to the Signature Generator which generates signature.

II. System Architecture

In this architecture we used a double honeypot system to detect new worms. Following figure 1 shows the system architecture of the system. Firstly, the incoming traffic goes through the Gate Translator which samples the unwanted inbound connections and redirects the samples connections to Honeynet1. The gate translator is configured with publicly-accessible addresses, which represent wanted services. Connections made to other addresses are considered unwanted and redirected to Honeynet 1 by the Gate Translator. Secondly, once Honeynet 1 is compromised, the worm will attempt to make outbound connections. Each honeynet is associated with an Internal Translator implemented in router that separates the honeynet from the rest of the network. The Internal Translator 1 intercepts all outbound connections from honeynet 1 and redirects them to honeynet 2 which does the same forming a loop. Only packets that make outbound connections are considered malicious, and hence the Double-honeynet forwards only packets that make outbound connections.

This policy is due to the fact that benign users do not try to make outbound connections if they are faced with non-existing addresses. Lastly, when enough instances of worm payloads are collected by Honeynet 1 and Honeynet 2, they are forwarded to the Signature Generator component which generates Signature. Signature generator consists of two honeypots, one high interaction, one low interaction and a Cloud AV which consist of ten antivirus engine and two behavioral detection engine. Here we are using sticky honeypot in between honeynet 1,2 and honeynet 3 to minimize instance of worm propagation and to generate effective signature for the worm using CloudAV. If cloudAV unable to detect worms then unused IP address system is automatically quarantined [6-7].Since honeypot 3 has set of blocks of antivirus to remove future polymorphic worms, which are developed with the help of behavioral detection engine which is deployed on unused system continuously till the removal of polymorphic worms.

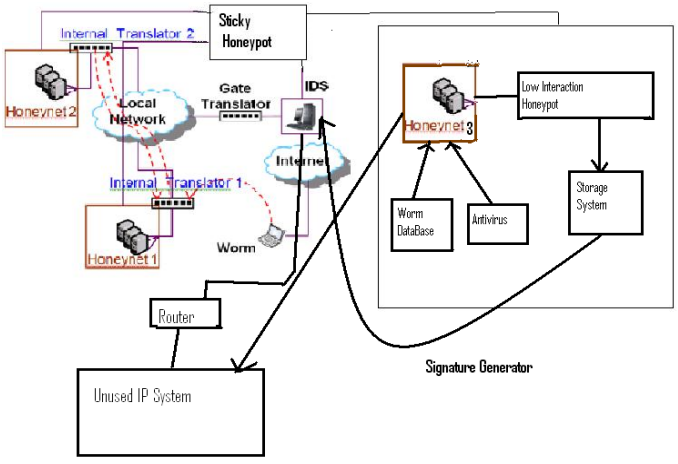


Fig I: System Architecture

III. Algorithm

a. Gate-translator collects incoming traffic and redirects them towards honeynet-1.
b. Internal translator implemented in router that separates honeynet from rest of the network.
c. Internal translator 1 intercepts all outbound connections from honeynet 1 and redirects them to honeynet 2.
d. When enough instances of worm payloads are collected by honeynet 1 and honeynet 2, they are forwarded to the signature generator.
e. Signature generator consist of two honeypots(one high interaction and one low interaction).When collected payloads are transferred to the Honeypot 3 ,we used sticky honeypot in between them that will minimize the worm propagation and signature will be generated at Honeypot 3.
f. Honeypot 3 has CloudAV antivirus which consist of ten antivirus engines and two behavioral detection engines which continuously run at Honeypot 3.
g. All the signatures are transferred to the storage system through low interaction honeypot. Then IDS can get all information about that payload.
h. If Cloud AV at Honeynet 3 unable to remove those worms then unused IP address system is automatically quarantined.
i. On quarantined unused system, blocks of future worm's removal capabilities antivirus is run continuously till it is removed.
j. After removal of polymorphic worm unused IP address system is again connected to the network.

IV. Conclusion

We have defined an algorithm to detect and defend newly detected polymorphic worms. The framework is designed using double honeynet and sticky honeypot. To detect newly polymorphic worms, we have used CloudAV antivirus which consist of ten antivirus engine and two behavioral detection engine, that continuously run at Honeynet 3.The undetected worms will be automatically quarantined at unused IP address system. In future we want to propose an automated signature generation system for polymorphic worms. We have proposed new detection method "Double-Honeypot" to detect new worms that have not been seen before. The proposed system will be based on Principal Component Analysis that will determine the most significant data that are shared between all 'polymorphic worms' instances and use them as signatures.

Reference

[1] L. Spitzner, "Honeypots: Tracking Hackers," Addison Wesley Pearson Education: Boston, 2002.

[2] Yong Tang, Shigang Chen," An Automated Signature-Based Approach against Polymorphic Internet Worms," IEEE Transaction on Parallel and Distributed Systems, pp. 879-892 July 2007.

[3] Snort – The de facto Standard for Intrusion Detection/Prevention, Available: http://www.snort.org, 14 February 2011.

[4] Bio Intrusion Detection System. Available: http://www.bro-ids.org/, 14 February 2011. International Journal for Information Security Research (IJISR), Volume 1, Issues 1/2, March/June 2011 Copyright

[5] John Oberheide.,Evan Cooke., Farnam .Jahanian., CloudAV: N Version Antivirus in the Network cloud, University of Michigan,Ann, Arhor, USENIX pp 1-18,2008.
[6] B.K.Mishra., N.Jha., SEIQRS model for the transmission of the malicious object in computer network, Applied Mathematical Modeling, 34, pp.710-715,2010.
[7] D.Moore.,C.Shannon., G.M.Valker., S.Savage., Internet Quarantine requirement for containing self replicating code ,Proceeding of the 22nd annual joint conference of the IEEE Computer and communication Societies, Infocom 2003, San Francisco, California, U.S.A, April, 2003.
[8] Cohen.F., Computer worms theory and experiment, Computer and Security, Vol 6,pp. 22-35,1987.
[9] Yong Tang and Shigang Chen., Defending Against Internet Worms: A Signature- Based Approach, Department of Computer & Information Science & Engineering, University of Florida, Gainesville, FL,USA., pp. 32611-6120,2010.
[10] Mohssen M. Z. E. Mohammed, H. Anthony Chan, Neco Ventura. "Honeycyber: Automated signature generation for zero-day polymorphic worms"; Proc.of the IEEE Military Communications Conference ,MILCOM, 2008.
[11] Tang,Y.; Chen, S. (2005). Defending Against Internet Worms: A Signature-Based Approach. In Proceedings of IEEE INFOCOM'2005, Miami, Florida, USA, pp.1-11.

Shape Prediction Linear Algorithm Using Fuzzy

Navjot Kaur[1]
Computer Science and Engineering
RIEIT
Ropar, India

Sheetal Kundra[2]
Computer Science and Engineering
RIEIT
Ropar, India

Harish Kundra[3]
Computer Science and Engineering
RIEIT
Ropar, India

***Abstract*— The goal of the proposed method is to develop shape prediction algorithm using fuzzy that is computationally fast and invariant. To predict the overlapping and joined shapes accurately, a method of shape prediction based on erosion and over segmentation is used to estimate values for dependent variables from previously unseen predictor values based on the variation in an underlying learning data set.**

Keywords- Shape prediction; Shape recognition; Feature extraction.

I. INTRODUCTION

Shape prediction represents an important domain of recognizing image objects, based on their shape information.[1] The Prediction of the original shape by efficient and accurate algorithm for connected or overlapping objects in an image leads to the decreased algorithm execution time and elapsed time. Each image contains up to several hundred objects, which were manually arranged not to overlap or touch each other. The algorithm is divided into two stages. In the first stage multiple thresh-holding values for the image are defined. Over segmentation and erosion is applied on binary image to erode away the boundaries of regions of foreground pixels.[9] And in the second stage features of the current object whose user is going to predict the shape are matched with the preloaded features in data set.[7] The equivalence distance to which the current object matched in data set is considered.

A. Morphology

The term Morphology refers to set of image processing operations that process images based on shapes. Morphological operations apply a structuring element to an input image, creating an output image of the same size. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbors. By choosing the size and shape of the neighborhood, you can construct a morphological operation that is sensitive to specific shapes in the input image.[2] Morphological operations affect the structure or shape of an object. All these operations are applied on binary images.

B. Structuring Element

Structuring element consists of matrix of 0's and 1's. Its size is smaller than the image and its origin identifies the pixel to be processed. The structuring element used for processing the images under prediction is disk shaped.[10]

If A and B be two sets in Z^2 then,

$$A \oplus B = \{ \mid (B)_z \cap A \neq \phi \}$$

Where A is image and B is the structural element.

C. Morphological Operations

The two principal morphological operations are dilation and erosion.

1) *Dilation*

Dilation allows objects to expand, potentially filling in small holes and connecting disjoint objects. Structural element of S is applied to all pixels of binary image. Every time the origin of the structural element is combined with a single binary pixel, the entire structural element is wrapped and subsequent alteration of the corresponding pixels of binary image.[3] The results of logical addition are written into the output binary image, which was originally initialized to zero.

$$A \oplus B = \{Z \mid [(B)_z \cap A] \varepsilon A \}$$

2) *Erosion*

Erosion shrinks objects by etching away (eroding) their boundaries. When using erosion structural element also passes through all pixels of the image.[4] If at a certain position every single pixel structuring element coincides with a single pixel binary image, then the logical disjunction of the central pixel structuring element with the corresponding pixel in the output image.

The method of erosion for prediction of overlapping and connecting images is specially used in this algorithm to increase the efficiency and improve execution time.

$$A \ominus B = \{Z \mid [(B)_z \varepsilon A]$$

Where A is an image and B is structuring element in Z^2.

D. Fuzzy Logic

A fuzzy system is represented by if-then rules in the form:

$$If\ i_1\ is\ vi_1,1\ and\ \cdots\ and\ i_m\ is\ vi_m,1$$

$$then\ o_1\ is\ vo_1,1\ and\ \cdots\ and\ o_n\ is\ vo_n,1$$

Where m is input and n is output, r is fuzzy rules in the system. The rules r defines the fuzzy rules which is an exponential function of the number of the inputs *i* and the number of linguistic values *l* taken by input.

$r = l^i$

If a fuzzy system has n inputs and single output then its fuzzy rules can be of the form:

$$If\ X_1\ is\ A_{1j}\ and\ X_2\ is\ A_{2j}\ \cdots\ and\ Xm\ is\ A_{jm}$$
$$then\ Y\ is\ B_j$$

E. Dataset

It is a collection of data elements. The following name/value pairs are used when a dataset is constructed:

1. VarNames: This gives the variables with the specified variable names.

{name_1,...,name_m}

2. ObsNames: This gives the n observations in A with the specified observation names.

{name_1,...,name_n}

II. Operational Stages

The techniques to estimate values for dependent variables from previously unseen predictor values based on the variation in an underlying learning database are used to predict the objects in the shape.

Main focus of work is to predict the shape by defining morphological operation which describes all boundary points of a shape. Development and Prediction of the shape by efficient, accurate, computationally fast and invariant algorithm for connected or overlapping objects in an image is the main consideration so that execution time and elapsed time is decreased.[6]

F. Basic Steps

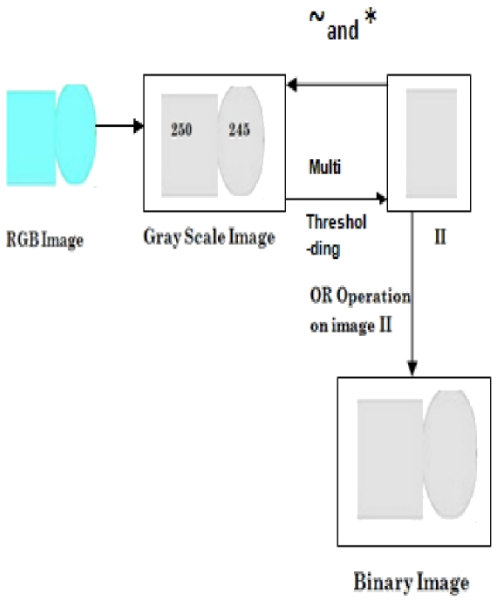


Figure1: Basic Steps Levels of algorithm are as under,

Level 1

1. Read Image: First step is to read the RGB image and convert that image to gray scale image by defining multi-thresholding.
2. Over segmentation: Next step is to do over-segmentation and convert the image to binary image.
3. Erosion: Apply erosion on binary image to erode away the boundaries of regions of foreground pixels (*i.e.* white pixels). Thus areas of foreground pixels shrink in size, and holes within those areas become larger.[5]
4. Feature finding: Find the features and edges for the current image that will be done with the help of fuzzy logic operations and will be loaded into memory for use whenever it is needed.

Level 2

The features of the current object whose user is going to predict the shape are matched with the preloaded features in data set. The equivalence distance to which the current object matched in our data set is considered.

III. Shape Prediction: The Algorithm

The detailed algorithm is described as,

1. Read an RGB image.

I= r(x,y), g(x,y), and b(x,y), a collection of image functions.

2. Convert an RGB image to a gray scale image.

J = rgb2gray (I)

Where J and I represents gray scale and rgb image

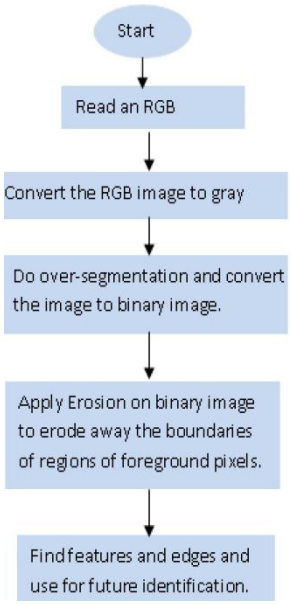


Figure2: Operating Stages for Level 1

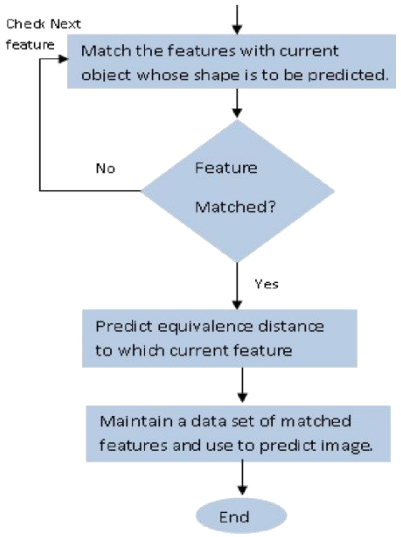


Figure3: Operating Stages for Level 2

3. Reduction to the information to make the processing of the visual data by defining threshold value.[8]

a) Defining Threshold value:

A new threshold is created that is the average of m_1 and m_2

$T' = (m_1 + m_2)/2$

where, m_1 *= average value of* G_1,

m_2 *= average value of* G_2 *and* G_1

$=\{f(m,n):f(m,n)>T\}$*(object pixels)*

$G_2 = \{f(m,n):f(m,n)<= T\}$ *(background pixels)*

f(m,n) is the value of the pixel located in the m^{th} *column,* n^{th} *row.*

b) Defining a structural element:

SE = strel(shape, parameters)

4. Apply Erosion on binary image.
5. Find features and edges.
6. Predict equivalence distance to which current feature matches. The minimum value among all predicted values is considered.
   $d = \sqrt{(x - a)^2 + (y - b)^2}$ $z = min(d)$

   where z is the minimum value and x represents the predicted values.
7. Establishment of the Prediction Table *prtbl* by using fuzzy method.
8. Use table *prtbl* for future identification and Match the features with current object whose shape is to be predicted.

Table1: Table prtbl for the predicted time values

Objects/Image	Predicted Time				
	Circle	*Triangle*	*Rectangle*	*Polygon*	*Eclipse*
10	5.77	5.22	5.28	6.18	5.43
13	9.60	8.85	8.36	9.16	8.54
13	7.86	7.21	7.22	8.18	7.50
15	8.76	8.32	8.19	9.52	8.24
15	9.13	8.60	8.62	9.72	8.75
15	8.62	8.09	8.01	21.76	8.18
16	9.96	9.13	9.19	18.86	9.47
17	11.18	9.99	9.94	11.40	10.35
17	10.54	18.64	9.59	26.01	9.99
22	13.94	29.49	23.20	33.50	13.24
692	60.35	190.99	185.02	189.41	199.66

The predicted time values Table1 above shows the predicted values for time which may vary depending upon input images and number of objects per image.

IV. RESULTS

For the purpose of testing the Shape Prediction algorithm, different input values are considered.

1. Output surface for circle:

Output surface of image with values for circle by taking Number of objects on x-axis and Time taken on the y-axis.

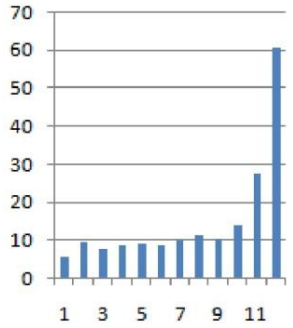


Figure4: Output surface for circle

This output surface shows the variation in the values of circle predicted in all the images. The predicted values vary as the number of objects in the image increases.

2. Output surface for eclipse:

Output surface of image with values for eclipse by taking Number of objects on x-axis and Time taken on the y-axis.

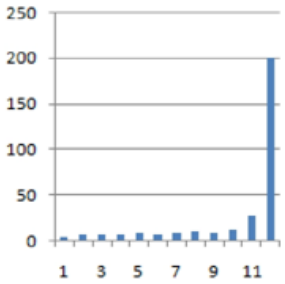


Figure5: Output surface for eclipse

This output surface shows the variation in the values of eclipse predicted in all the images. The predicted values vary as the number of objects in the image increases.

3. Output surface for polygon:

Output surface of image with values for polygon by taking Number of objects on x-axis and Time taken on the y-axis.

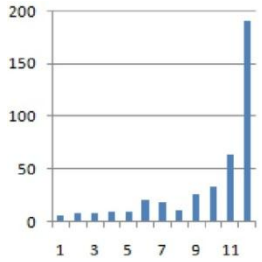


Figure6: Output surface for polygon

This output surface shows the variation in the values of polygon predicted in all the images. The predicted values vary as the number of objects in the image increases.

4. Output surface for rectangle:

Output surface of image with values for rectangle by taking Number of objects on x-axis and Time taken on the y-axis.

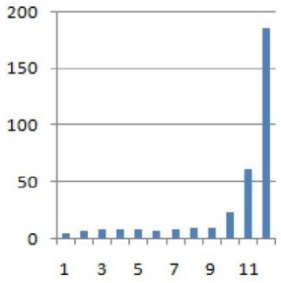


Figure7: Output surface for rectangle

This output surface shows the variation in the values of rectangle predicted in all the images. The predicted values vary as the number of objects in the image increases.

5. Output surface for triangle:

Output surface of image with values for rectangle by taking Number of objects on x-axis and Time taken on the y-axis.

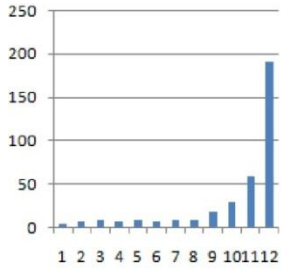


Figure8: Output surface for triangle

This output surface shows the variation in the values of triangle predicted in all the images. The predicted values vary as the number of objects in the image increases.

6. Combined output surface for values of Circle, Eclipse, Polygon, Rectangle and Triangle:

The number of objects per image can be any; here the considered objects per image are 5.

Output surface of image with values for all objects as combined image by taking number of objects on x-axis and time taken on the y-axis.

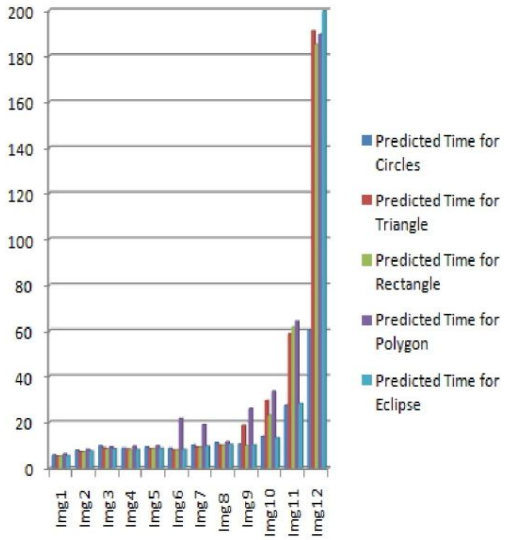


Figure 9: time taken and number of objects recognized

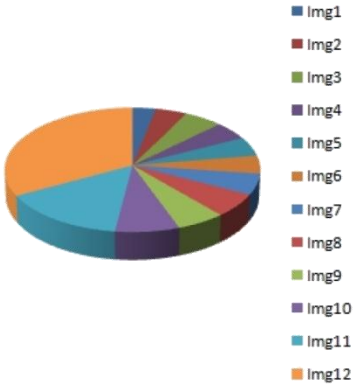


Figure10: Pie chart representation for Combined output surface

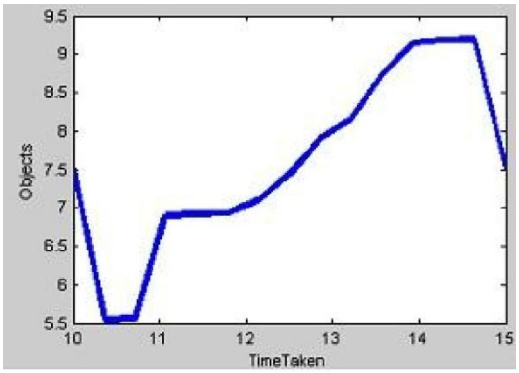


Figure11: Output surface of time taken and number of objects recognized by defining fuzzy rules

V. EXPERIMENTAL RESULT

Technique used to extract characteristics in order to get the prediction characteristic of different shapes. 12 samples are extracted randomly for each of the five kinds of shape to work out the statistical distribution rule of the characteristic value. The image quality measures like PSNR, MSE etc. are used to find the correct recognition values for objects of image and at last the predicted values used to get the recognizable results, as shown in table 2. It can be seen that the correct recognizable rates of eclipse, triangular, rectangle and circular shapes all exceed 90%, and the total correct recognizable rate amount to 94.5%. Table2 gives the recognizable results of shape prediction based on fuzzy method.

Table2 The recognizable results of shape prediction based on fuzzy method

Correct recognition number	Correct recognition rate (%)	Total recognition rate (%)
4.9	95	94.50
4.7	95	
4.8	95	
7.1	93	

VI. FUTURE SCOPE

In this paper, the shape prediction method based on fuzzy is proposed. This method efficiently recognizes the different individual and overlapping objects. The utilization of proposed method could be in robotics, in which robots are used to predict the objects which could remain unpredicted from human. Image definition can be done to define the image based on characteristics of objects predicted.

This recognizable prediction method based on fuzzy can reflect different shapes objectively and correctly, so the total automatically recognizable rate arrives at 94.50%.

REFERENCES

[1] Wang, R., Bu, F., Jin H., and Li, L., 2007. "Toe shape recognition algorithm based on fuzzy neural networks", IEEE International conference on Natural Computation", vol. 2, pp. 737-741.

[2] Wang, S., Wang, Y., Gu, X., and Samaras, D., 2007. "Conformal Geometry and Its Applications on 3D Shape Matching, Recognition and Stitching", IEEE International conference on control and design, vol. 29, issue 7, pp. 1209 1220.

[3] Mingqiang, Y., Kidiyo, K., Joseph, R., 2008. "Shape Matching and Object Recognition Using Chord Contexts", IEEE conference on Object recognition, vol. 4, pp. 63-69.

[4] Baloch, S. and Krim, H., 2010. "Object Recognition through Topo-Geometric Shape Models Using Error-Tolerant Subgraph Isomorphisms", IEEE Transactions On Image Processing, vol. 19, issue. 5, pp. 1191-1200.

[5] Salve, G., Jondhale, C., 2010. "Shape Matching and Object Recognition Using Shape Contexts", IEEE conference, vol. 9, pp. 471-474.

[6] Song, R., Zhao, Z., Li,Y., Zhang, Q., Chen,X., 2010. "The Method of Shape Recognition Based on V-system", IEEE Conference on Frontier of Computer Science and Technology, fcst, pp. 321-326.

[7] Yu, D., Jin1, S., Luo1, S., Lai, W., Park, M and Pham, T., 2010. "Shape Analysis and Recognition Based on Skeleton and Morphological", International Conference on Computer Graphics, Imaging and Visualization, cgiv, pp. 118-123.

[8] Almazan, J., Fornes, A., Valveny, E., 2011. "A Non-Rigid Feature Extraction Method for Shape Recognition", International Conference on Document Analysis and Recognition, issue 4, pp. 987-991. Barbu, T., 2011. "Automatic Unsupervised Shape Recognition Technique using Moment Invariants", IEEE conference, pp. 1-4.

[9] Yuan, W., Jing, L. 2011. "Hand-Shape Feature Selection and Recognition Performance Analysis", IEEE conference, ichb, pp. 1-6.

AUTHORS PROFILE

Er. Navjot Kaur received the B.Tech degree in computer science and engineering from Punjab Technical University, India. She is perusing her master's degree in computer science and engineering from Punjab Technical University.

Er. Sheetal Kundra did her M.Tech degree in computer science and engineering from Punjab Technical University, India. She has been working as an assistant professor in computer science department at RIEIT, Ropar, India.

Er. Harish Kundra is an Associate Professor and Head in the Department of Computer Science & IT at Rayat Institute of Engineering and IT, Near Rupnagar, Punjab since 2002. He did his BE in Computer Science & Engineering in 2000 and M-Tech in 2007. He is pursuing his Ph.D Degree in the area of Artificial Intelligence from Punjab Technical University, Jalandhar. He has published over 80 papers in various National and International journals/conferences. He has supervised 40 UG Major Projects, 15 M.Tech. Dissertations.

The Development of Mobile Client Application in Yogyakarta Tourism and Culinary Information System Based on Social Media Integration

Novrian Fajar Hidayat
Department of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia

Ridi Ferdiana
Department of Electrical Engineering and Information Technology
Universitas Gadjah Mada
Yogyakarta, Indonesia

***Abstract*— Social network is currently being an important part of someone. Many of users in social network make it an effective publication. One of many things that can be published on social network is tourism. Indonesia has a lot of tourism and culinary, especially on Special District of Yogyakarta.Tourism and culinary resources on Yogyakarta can be published and shared using social network. In addition, development of mobile technology and smartphone make easier to access social network through internet.**

The release of Windows Phone 7 makes new color in the world of smartphone. Windows Phone 7comes with elegant interface, Metro Style. Besides that, standardized specification makes Windows Phone 7 suitable for integrating social network with tourism and culinary on Special District of Yogyakarta.

This Research is expected to integrate social network with tourism and culinary on Yogyakarta. The method in this research is using ICONIX method. This method is one method that combines waterfall and agile methods. The results of this study are in the form of applications that run on Windows Phone 7 and consume a web service. This application provides information especially for tourist in order to be able to easily find culinary and tourism in Yogyakarta.

Keywords- Social Network; Information Service; Culinary; Tourism; Windows Phone 7.

I. INTRODUCTION

Nowadays, 200 milion users login daily into Facebook. In the same time, 95 milion tweets are written[11]. This fact are prove of many users of social network. These resource are on the social network to take the advantage especialy for tourism.

Special District of Yogyakarta, known as region that have a lot of tourism resource. In other side, the development of technology makes internet gives many varies of information. Social network also become important and give contribution for information on internet.

Mobile Client Application provides application service on mobile devices. It is different with web client application, Mobile Client Application could use device specific features such as GPS, camera, and sensors.

One of a smartphone that use mobile client application model is Windows Phone 7. IDC said that Windows Phone 7 would have good sales potential after Nokia introduced the product that using Windows Phone 7 operating system. It would get 20% market of smartphone [7]. Furthermore, Developing Windows Phone 7 application still has huge potential market since the market application is not huge as Android or iPhone.

Another advantages of developing Windows Phone 7 based aplication are developer no need to worry about the compatibility from one to another device. Windows Phone 7 have standardized hardware specification in order to execute to different device. With Windows Phone 7, integrated application between tourism and social network on Special District of Yogyakarta could be developed. Furthermore, users could share infomation on their social network about tourism or culinary that they liked.

II. BASIC THEORY AND LITERATURE

A. Basic Theory

1) Social Network

Social Network sites like Facebook or Myspace are network with many users could add detail profile, so they could communicate with others [8]. Social Network is part of social media that focusing in sharing contents or activities. Nowadays, 200 million users login to Facebook every day and in the same time, 95 tweets written every day. Half of the users using mobile device to do it [11].Therefore, the connection any system with social network shall help the application to get the benefit in social media.

2) Tourism Information System

Tourist Information System is a service that provided by an organization for giving information to tourist about attraction, events, or culinary. It could be web or mobile client. Users could use it for tourism information guide when visiting a point of interest [5].

3) Mobile Client Application

Mobile Client defined by Darren Ince[3] is a client that not only on one position: in example, a portable computer using mobile phone to communicate with server or a mobile phone use wireless application protocol to connect on web server. Mobile client application runs on client side to provide the rich

experience to the client through the benefit of the existing mobile application platform.

4) Windows Phone 7

Microsoft with project codenamed 'Metro' on February 2010 to make new mobile operating system [6]. This operating system known as Windows Phone 7. It adopted Metro interface that different with former Windows Mobile and also another smartphone like Android or Apple iPhone. Windows Phone UI focuses in simplicity and consistency of user experience.

B. Previous Researches

Kurniawan [10] researched about Mobile Client Application and tourism in Yogyakarta with developing application based on J2ME. This research use RSS 2.0 for providing information into mobile device. It provides read only information about tourism.

Another research about tourism and culinary information service on Yogyakarta also already been done in mobile web platform. It is used mobile web application to displaying information on mobile device [9]. This research already uses social network and flexible user interface.

The former research used Android Smartphone to running an application that provides user information about tourism and culinary [1]. This application also could access GPS and share to social network. This application named 'JOGJANAN'. Based on the previous researches, Table II shows the differentiation between this application with the previous researcher

TABLE I. COMPARISON WITH SIMILAR APPLICATION

Author	***Platform***	***Benefits***
Arfian, 2010	Mobile Web	Adaptive web can be used in various mobile devicesbased on HTML and CSS standard
Kurniawan, 2010	J2ME Client	Focusing in fast and lightweight application based on RSS technology
Febriyanto, 2011	Android	Focusing in smartphone android platform based on Web Services technology
This Research	Windows Phone	Focusing in smartphone, windows phone platform uses REST and Social Media

III. PREPARERESEARCH METHOD

When developing this software, researcher adopted ICONIX method. ICONIX method is method that combine classical or waterfall method and agile method [4]. It is start from make use case diagram, domain modeling diagram, robustness diagram, sequence diagram, and the last static class diagram. ICONIX is chosen because the development of Jogjanan mobile application needs lightweight software engineering method. ICONIX is already familiar with the technical team who build the software. Figure 1, shows the ICONIX method approach.

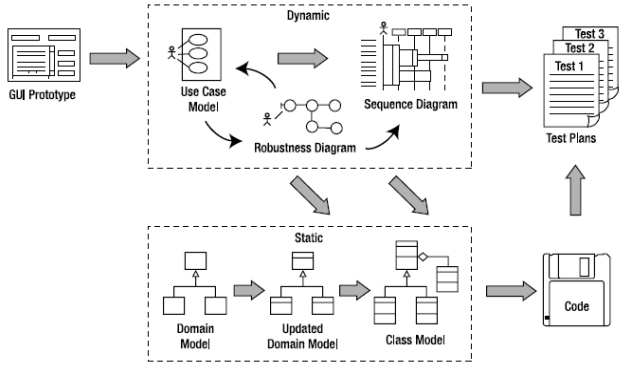


Figure 1. ICONIX Method[4]

IV. RESULTAND REVIEW

A. Application Architecture

This application uses MVVM (Model View View Model) design pattern. This design pattern fit to use on an application that implement mark-up language such as XAML.

View is part that directly communicated with user. It is also contain user interface. View Model is the logic part and bridge between View and Model. Model is representing from object data.

In this architecture, researcher added Helper as a component that could not include to Model, View or View Model. Figure 2, shows the application architecture of Jogjanan. It is used layered architecture that is combined with MVVM model.

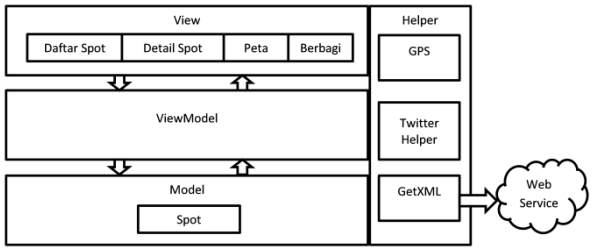


Figure 2. Application Architecture

Above the application architecture, this application connected with system that developed from former research. It had been developed web service as data center for consumed by this application. Figure 3, shows the system architecture of Jogjanan mobile application.

B. Web Service Testing

This application is a part of tourism and culinary information system that have been developed. Therefore, the web services are already developed by the previous researches [1] [10].

In order to get the data from a server, mobile client has an access to provide web service. Web service is provided through WSDL (web services development language). The web service is developed by PHP and nuSOAP library.

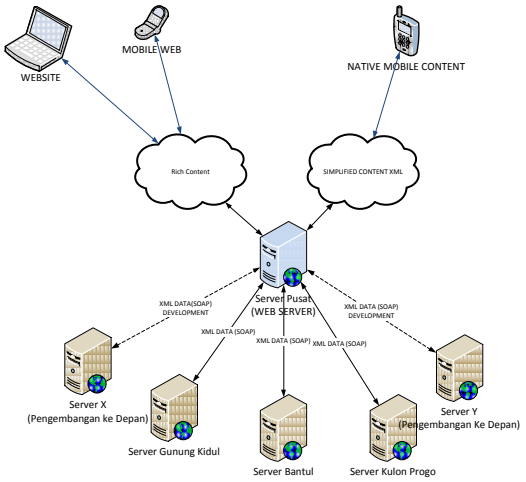


Figure 3. System Architecture [2]

However, during the testing mode, the web services couldn't be connected to Windows Phone. Researcher does several tests and shows that that the WSDL format is not suitable with Windows Phone.

Therefore, the web service is extended by providing a REST result (Representational State Transfer). This web service will give output as XML data. The XML data is consumed through Windows Phone user interface

C. Spot List Testing

When application running, user will enter main page, Spot List. This application will call web service that located on Jogjanan service through SOAP Header authentication model. After application got the data, it will displays on the screen as culinary and spot using two ways binding technique. Two ways binding provides side by side binding between client and the web services. Figure 4 shows the binding result in user interface.

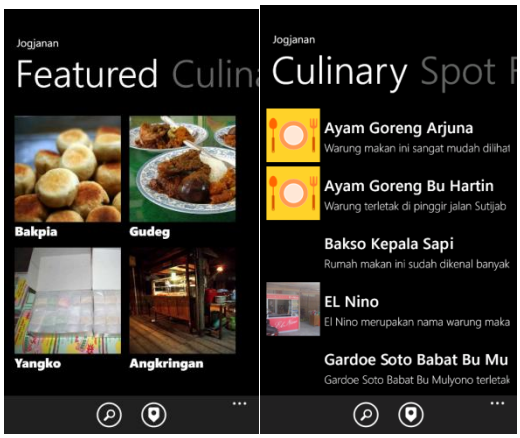


Figure 4. Featured page (left) and Culinary page (right).

D. Detail Spot Testing

When user click or tap at a content on the list, the application will navigate to detail spot page.

This page displayed spot name, image thumbnail, spot description, facilities, a button to navigate to map and share button.

This feature uses progressive and scaling image binding to display image in sufficient resolution. Figure 5 shows the spot detail in user interface.

Figure 5. Detail Page

E. Social Network Sharing Testing

This testing uses twitter as social network for sharing. After tap or click share button on detail spot then user will ask username and twitter password for authentication. Then, the user could write a message that will be published on Twitter. It is automatically include the URL of spot. It will share the nice URL to the user peers in social network. Figure 6 shows the twitter display when share the Jogjanan POI (point of interest).

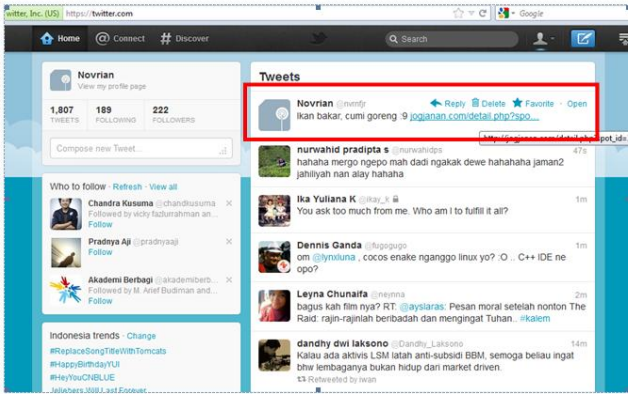


Figure 6. Twitter Display

F. Map Testing

Map feature could be accessed by tapingthe map icon on Spot list page. Map feature will display the location of tourism and culinary spot. Besidesof that, the map feature display user location by using A-GPS that already integrated in windows Phone devices. This map can zoom in or out as shown in Figure 7.

Figure 7. Map Page

G. Marketplace Verification and Validation

Visual Studio 2010 for developing Windows Phone 7 application provide testing tool in order to check an application. There are four items that are verified on Visual Studio which are packaging format, validation user experience, iconography standard, and application screenshot.

In this test, JogjananPhone application is verified and validated on the Visual Studio Windows Phone Testing kit. Figure 8 shows the Testing toolkit helps validation and verification for market store purpose.

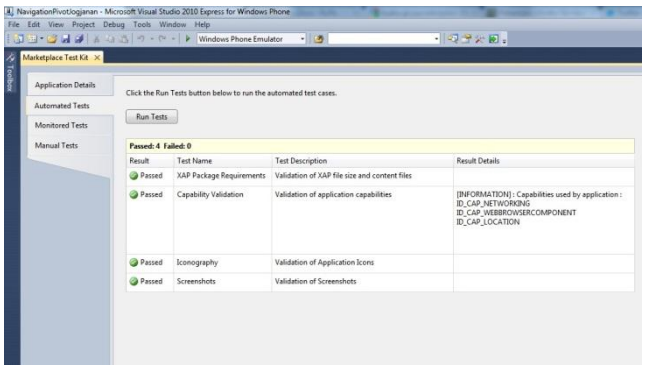


Figure 8. Result of Marketplace Verification and Validation test

H. Comparison with The Existing Application

TABLE II. COMPARISON WITH SIMILAR APPLICATION

Comparison	***Toresto***	***JogjananPhone***
Application Focus	Culinary Information system	Tourism and Culinary Information System
Share to social network	Share to Facebook and Twitter	Can share to Facebook and Twitter
Tourism object	Not providing tourism object	Provide and displaying tourism object
Maps	On the detail culinary page	Different page
Programming Language	Java	.NET
Maps Type	Google Map	Bing Map
Point of interest	More than 100 POI	60 POI
Contextual POI (Nearby)	Yes using GPS	Yes using GPS

Toresto is similar application that running in Android operating system is already on Google Play since 2011. Toresto focuses in culinary system, meanwhile the Jogjanan focuses in culinary and tourism system. Just like Toresto, Jogjanan also available in others platform such as web, mobile, and rich client.

Table II shows the differentiation between this applications with the existing application that exist in Windows Phone 7 marketplace. By seeing the table, is concluded that both application has a differentiation in term of user experience and functionality.

V. CONCLUSION AND FUTURE WORKS

After analyzing, testing, and reviewing, it could concluded that this research provides several conclusion which are.

- The mobile application that is developed in this research provides a good insight that many of mobile application use social media as a tool to increase the attractiveness in the application
- The geo-location, maps, and social media combines the unique values of the mobile application. In this research, it is shown that these items provide a "contextual based experience"
- The mobile application should be more accessible for tourist since the existence of application store. However, it made several testing and deployment application harder

This application is just like others mobile application need to be refined in term of user experience. Therefore, it needs a further research about user experience evaluation for contextual based experience.

Several agenda that can be done in the future also related with the need of the application to evaluate the usage effectiveness with or without social media.

VI. REFERENCE

[1] A. Febriyanto. 2011. Perancangan Aplikasi Android untuk Layanan Informasi Wisata dan Kuliner Daerah Istimewa Yogyakarta. Yogyakarta: Electrical Engineering and Information Technology, Gadjah Mada University.

[2] A. Saktiaji. 2011. Perancangan Web Service Sebagai Sarana Komunikasi Layanan Informasi Wisata dan Kuliner di Daerah Istimewa Yogyakarta berbasis NuSOAP. Yogyakarta: Jurusan Teknik Elektro dan Teknologi Informasi, Universitas Gadjah Mada.

[3] D. Ince. 2001. A Dictionary of the Internet. http://www.encyclopedia.com. accessed 19 October 2011.

[4] D. Rosenberg,M. Stephens & M. Collins-Cope. 2005. Agile Development with ICONIX Process: People, Process, and Pragmatism. New York: Apress.

[5] H.Afsarmanesh&L. M. Camarinha-Matos, 2000. Future Smart-Organizations: A Virtual Tourism Enterprise. International Conference on Web Information System Engineering (WISE'00).

[6] H. Lee& E.Chuvyrov. 2011. Beginning Windows Phone 7 Development Second Edition. New York: Apress.

[7] K. Nagamine. 2011. Worldwide Smartphone Market Expected to Grow 55% in 2011 and Approach Shipments of One Billion in 2015, According to IDC. http://www.idc.com/getdoc.jsp?containerId=prUS22871611. accessed 19 October 2011.

[8] P.Walden &C. Carlsson. 2011. Travel Information Search – The Presence of Social Media. Proceedings of the 44th Hawaii International Conference on System Sciences.

[9] R. A.Arfian. 2010. Pengembangan Aplikasi Mobile Web untuk Wisata dan Kuliner Yogyakarta. Yogyakarta: Jurusan Teknik Elektro dan Teknologi Informasi, Fakultas Teknik, UGM.

[10] R. E. Kurniawan. 2010. Perancangan aplikasi Client Mobile Web Service Info Tempat Wisata di Yogyakarta dengan J2ME dan RSS 2.0. Yogyakarta: Jurusan Teknik Informatika, STMIK AMIKOM.

[11] S. Dixon. 2011. Social: Holy Crap Social Media Statistics 2011. http://www.ecademy.com/node.php?id=166989.accessed 19 October 2011.

AUTHORS PROFILE

Novrian Fajar Hidayat, is an engineer from Universitas Gadjah Mada. He focuses in User Experience engineer and creative media development. In his daily life he joins Microsoft Innovation Center as a UX Specaialist. He can be reached at thenovrianmail@gmail.com

Ridi is a lecturer and researcher in Universitas Gadjah Mada. He finished doctoral degree in Software Engineering (ALM) focusing in Agile methodology. He has several Microsoft certifications such as MCTS, MCPD, MCITP, and MCT. Nowaday, Ridi's loves to write his tought at blog.ridilabs.net or on his twitter at @ridife

An Empirical Analysis Over the Four Different Feature-Based Face and Iris Biometric Recognition Techniques

Deepak Sharma
Department of CSE & IT
Kurukshetra Institute of Technology & Management
Kurukshetra, Haryana, India

Dr. Ashok Kumar
Department of Computer Science Applications
Maharishi Markandeshwar University (MMU)
Mullana, Ambala, Haryana, India

***Abstract*— Recently, multimodal biometric systems have been widely accepted, which has shown increased accuracy and population coverage, while reducing vulnerability to spoofing. The main feature to multimodal biometrics is the amalgamation of different biometric modality data at the feature extraction, matching score, or decision levels. Recently, a lot of works are presented in the literature for multi-modal biometric recognition. In this paper, we have presented comparative analysis of four different feature extraction approaches, such as LBP, LGXP, EMD and PCA. The main steps involved in such four approaches are: 1) Feature extraction from face image, 2) Feature extraction from iris image and 3) Fusion of face and iris features. The performance of the feature extraction methods in multi-modal recognition is analyzed using FMR and FNMR to study the recognition behavior of these approaches. Then, an extensive analysis is carried out to find the effectiveness of different approaches using two different databases. The experimental results show the equal error rate of different feature extraction approaches in multi-modal biometric recognition. From the ROC curve plotted, the performance of the LBP and LGXP method is better compared to PCA-based technique.**

Keywords- Multi-modal biometrics; Face Recognition; iris recognition; LBP operator (Local Binary Pattern); Local Gabor XOR Patterns; PCA and EMD.

I. INTRODUCTION

Over the past decade, biometric authentication has drawn substantial attention with growing demands in automated personal identification [4]. This is due to the reason that traditional automatic personal identification systems use tools such as Personal Identification Number (PIN), ID card, key, etc., to verify the identity of a person. But in modern world, such tools are not reliable enough to fulfill the security necessity of person authentication system [5]. A biometric system provides automatic identification of a person by considering some unique features or traits obtained by the person [6]. Some common biometric features are fingerprints, hand-geometry, face, voice, iris, retina, gait, signature, palm-print, ear [5]. An excellent biometric can be identified by use of a feature i.e., the feature should be highly unique - so that the chance of any two person containing the same feature will be negligible, stable – so that the trait does not differ over time, and be easily captured – in order to provide expediency to the user, and prevent falsification of the feature [6]. Multimodal biometric systems are capable of defeating some of the restrictions of unimodal biometric systems by adding multiple sources of information for the process of personal recognition [7]. Such systems are expected to be more dependable because of the existence of several, autonomous fragments of evidence [8, 9].

Face recognition is one of the most commonly used biometric features [11]. The intent of face recognition is to identify the persons in images or videos from their facial expression. When a comparison is done with other biometrics, it is found that the face recognition is passive and does not necessitate cooperative persons who are close to or in contact with a sensor [10]. The face is an easily collectible, universal and non-intrusive biometric [12], which makes it perfect for applications while other biometrics such as fingerprints or iris scanning are not possible.

Among the present biometric features, iris is one of the most reliable and precise biometric trait. Iris recognition is widely recognized as one of the preeminent biometrics recognition techniques in the world due to its constancy, distinctiveness, and non-invasiveness, which also has the potential of applications in wide areas [13] [14]. Iris is an externally visible, yet protected organ whose unique epigenetic pattern stays constant throughout the adult life [15]. These features make it suitable for use as a biometric for identifying the persons. Image processing methods are utilized to get the unique iris pattern from a digitized image of the eye, and encrypt it into a biometric template, which can be stored in a database [16]. This biometric template has an objective mathematical representation of the unique data stored in the iris, and permits comparisons to be done among the templates [15]. In iris recognition system, for identifying a person, initially his/her eye is photographed, and then a template is created for their iris region [16]. This template is then compared with the other templates stored in a database. If the matching template is found, the subject is identified or else, the subject remains unidentified [17].

In unimodal biometric systems, only a single feature is employed for validating the identity of a person and such systems are mostly affected by numerous practical problems such as noisy sensor data, non-universality or lack of uniqueness of the chosen biometric feature, unacceptable error rates, and spoof attacks [18]. Multimodal biometric systems

conquer most of these drawbacks by combining the proofs obtained from various sources [20] [19]. Multimodal biometrics has produced better accuracy [22] and population coverage, while reducing susceptibility to spoofing. The major feature to multimodal biometrics is the amalgamation of various biometric modality data at the feature extraction, matching score, or decision levels [21]. Numerous multimodal biometrics techniques and algorithms have been proposed by several researchers [23-30].

In our previous work [1], we have made use of LGXP feature as for face and iris-based multi-modal biometric recognition system. In this work, we have done a detailed analysis over our previous technique [1] with the technique given by Zhifang Wang *et al.* [2] who combined iris and face features for multi modal biometric recognition system using PCA and Gabor feature. The analysis is conducted over the two different set of face and iris image databases taken from CASIA and AT&T. Here, the performances of the previous technique with LGXP-feature, LBP-based feature, EMD (Empirical Mode Decomposition) feature and the technique proposed by Zhifang Wang *et al.* [2] are extensively analyzed with the help of FMR and FNMR curve. Finally, ROC graphs in between the FNMR vs. FMR that are the two chief metrics used in this experimentation is drawn to demonstrate the effectiveness of the previous approach. The rest of the paper is organized as follows: Section 2 describes the description of the methods taken for analysis for iris and face recognition. Section 3 discusses about the results discussion and section 4 concludes the paper.

II. Description Of The Methods Taken For Analysis

In the proposed approach, we compare the four feature extraction methods such as, Local Binary Pattern (LBP), local Gabor XOR pattern (LGXP), Empirical Mode Decomposition (EMD) and Principal Component Analysis (PCA) in face and iris-based multi modal recognition. The overall block diagram of our comparative analysis is shown in fig 1.

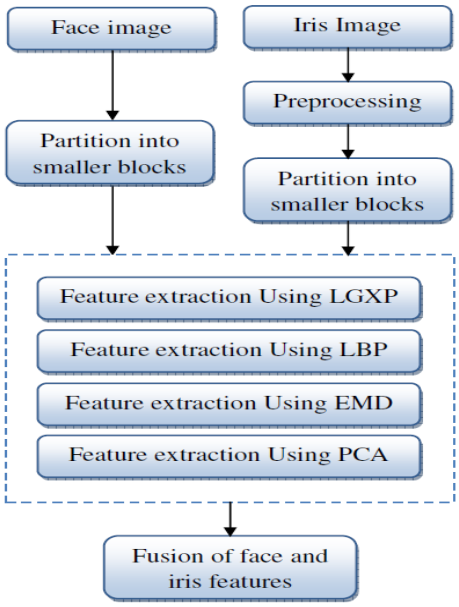


Fig.1: The Overall block diagram

A. LGXP-feature-based face and iris recognition system

In the first method, we make use of LGXP feature for multi-modal biometric recognition system. It consists of three modules such as, 1) Preprocessing 2) Feature Extraction of face and iris images and 3) Feature Matching.

1) Preprocessing

In the preprocessing stage**,** initially the iris image is converted into its normalized form via some preprocessing techniques before the feature extraction is made. The normalized form of the iris image is often utilized by the researchers to extract the features for iris recognition. Then, the iris segmentation is performed for identifying the iris boundary. Because, detecting the inner and outer boundaries of the iris texture is very crucial for effectual feature extraction. Integro-differential, Hough transform, and active contour model are some of the robust methods employed for detecting the boundaries. Eventually, a normalized iris image is obtained from the preprocessing step.

2) Feature Extraction of face and iris images:

In the second module, the normalized iris image and the facial image are partitioned into many small blocks and the pixel values in each block are converted into vector. Then, $1-U$ vector is applied to linear scaling and LGXP, which provides the feature vector. The irrelevant pixels (i.e., pixels in the eyelash and eyelid areas) in case of iris are calculated on each block for knowing the block importance. At last, we concatenate the feature vector of both the iris and face images and also the count of immaterial pixels of the iris image. The following steps depict the entire procedure of feature extraction of face and iris images:

- Partition the normalized iris image and the input face image into several small blocks.
- Convert the block of size $d_1 \times d_2$ into $1-U$ vector.
- Perform LGXP on each block of both the normalized iris image and the input face image. In this stage, we apply LGXP to the normalized iris image as well as to the face image. LGXP is applied on each of the rescaled 1-U vector D_V^N in order to acquire the feature of block D. After all the iterations, the LGXP creates a residual vector for both the iris and face images. The feature vector F^q and I^q (residual) of the block in face image and the iris image respectively, are represented as follows:

$$F^q = (F_1^q, F_2^q, \ldots\ldots\ldots\ldots, F_m^q)$$

$$I^q = (I_1^q, I_2^q, \ldots\ldots\ldots\ldots\ldots, I_m^q)$$

Where, F^q *and* I^q represents the residual of the LGXP results of D_V^N.

- Combine the feature vector & the count of immaterial pixels of the iris image and the feature vector of the face image.
- The resultant concatenated feature vector is given as,

$$T^q = (F_1^q, F_2^q, \ldots\ldots\ldots\ldots, F_m^q, I_1^q, I_2^q, \ldots\ldots\ldots\ldots\ldots, I_m^q, C_r))$$

Where, F^q *and* I^q represents the residual of the LGXP results of D_V^N for the face and iris image respectively and C_r is the number of immaterial pixels for a block D.

3) Feature Matching:

The subsequent steps describe the process involved in the matching phase.

Calculation of Weightage Based on Irrelevant Pixels: The concatenated feature vector is computed for a test sample having both the iris and face images, and it is compared with the concatenated vectors of an iris image and face image of the database. For each block, we compute the block weightage based on the irrelevant pixels of test sample and an iris and face images of database.

Let the number of irrelevant pixels corresponding to the first block of a test sample be C_s and the image of database be C_r. Then, the weightage W_y of the first block is calculated using the formula given below,

$$W_y = \begin{cases} 1 - \dfrac{C_y}{d_1 \times d_2} \; ; & if \; \dfrac{C_y}{d_1 \times d_2} < 0.5 \\ 0 \; ; & otherwise \end{cases}$$

Where, $C_y = \max imum(C_r - C_s)$

Using the above equation, the blocks are obtained with three diverse set of values. $W_y = 1$ represents the block without noise, $0.5 < W_y < 1$ represents the block with partial noise, and $W_y = 0$ represents the block with noise. Likewise, we compute the block weightage for all the blocks with respect to test image and sample of the database.

Score Computation Using Distance Metrics: Here**,** Euclidean distance (ED) measure is adopted for the score computation. Selecting a suitable similarity measure for matching the feature vectors is vital and selection of distance measure compliments the proposed technique. This metric provides a score, which represents the similarity between two feature vectors (i.e., test sample and images from database). Therefore, for each block, the ED measure S_x between the feature vectors say T_r^q and T_t^q is determined by using the following equation,

$$S_x = ED(T_r^q, T_t^q) = \sqrt{\sum_{i=1}^{m} (T_{r_i}^q - T_{t_i}^q)^2}$$

Average Matching Score Based on Weightage: In the prior steps, we have obtained a block weightage and matching score for all the blocks. These two values are then exploited to calculate the average matching score of the iris image. The formula used for computing the average matching score M_{avg} is given below,

$$M_{avg} = \frac{\sum_{b=1}^{n} M_b}{\sum_{b=1}^{n} (1 - W_{y_b})}$$

Where, $M = S_x \times (1 - W_y)$

$M \rightarrow$Weightage of block pair.

After performing the average matching of all the images, we obtain the value (M_{avg}). Then, this value is employed to determine whether the test sample is already present in the database or not. If the obtained value M_{avg} is below the predefined threshold (P), then we can decide that the test sample is present in the database.

B. LBP-feature-based face and iris recognition system

In the second method, we make use of LBP feature for multi-modal biometric recognition system instead of LGXP feature. It consists of three modules such as, 1) Preprocessing 2) Feature Extraction of face and iris images and 3) Feature Matching.

i) Preprocessing: Initially, the input iris and face image is transformed to its normalized form. The above mentioned process is also used here to transform it to the normalized form.

ii) Feature Extraction of face and iris images**:** The whole procedure of feature extraction of face and iris images includes following steps: 1) Partitioning of Normalized Iris Image, 2) Conversion of Block of Size $d_1 \times d_2$ into $1-U$ Vector, 3) Performing LBP on Each Block of both the Normalized Iris Image and the Input Face Image. 4) Concatenation of Feature Vector and the Count of Irrelevant Pixels of the Iris Image and the Feature Vector of the Face Image.

LBP operator: Local binary patterns were introduced by Ojala *et al.* [31] as a fine scale texture descriptor. In its simplest form, an LBP description of a pixel is created by thresholding the values of the 3x3 neighborhood of the pixel against the central pixel and interpreting the result as a binary number. Figure 2 illustrates the basic LBP operator.

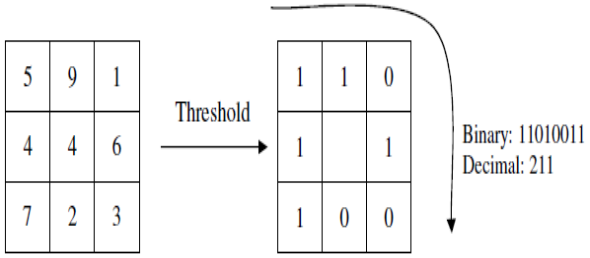


Fig.2: The basic LBP operator

3) Feature Matching: In feature matching, Weightage is computed for every block using the weightage equation defined above and matching Score is computed using the following formulae.

$$M_{avg} = \frac{\sum_{b=1}^{n} M_b}{\sum_{b=1}^{n} (1 - W_{y_b})}$$

Where, $M = S_x \times (1 - W_y)$

$M \rightarrow$ Weightage of block pair.

We obtain the value (M_{avg}) after performing average matching of all the images. This value is used to decide whether the test sample is already present in the database or not. If the obtained value M_{avg} is less than the predefined threshold (P), then the test sample is present in the database.

C. EMD feature-based face and iris recognition system

Here, we make use of Empirical Mode Decomposition (EMD) as feature for performing the multi-modal biometric recognition system. The remaining process (preprocessing and feature matching) is exactly same as that of previous step described in above sections.

Empirical Mode Decomposition: For the time-frequency analysis of real-world signals, an algorithm called empirical mode decomposition (EMD) is devised in [32]. By using EMD, the signal in hand is decomposed into numerous oscillatory modes called intrinsic mode functions (IMFs) [22, 26, 43]. *Intrinsic Mode Functions* (IMF): A function is an IMF, if and only if it satisfies two conditions: *Condition 1:* The number of extrema and the number of zero crossings differ by no more than one. *Condition 2:* At any point, the average of the envelope defined by the local maxima and by the local minima is zero. To decompose a signal by means of EMD, a sifting process [44] is used to extract the IMF from a given signal. *Sifting Process:* The procedure to decompose a signal $S(t)$, where $t = \{1,2,\cdots,T\}$ and T is the total no. of samples in $S(t)$, is explained in detail as follows:

Step 1: Determine all the maxima points in the given signal $S(t)$ and generate the upper envelope say $S_{up}(t)$ via Cubic Spline Interpolation.

Step 2: Determine all the minima points in the given signal $S(t)$ and generate the lower envelope say $S_{low}(t)$ via Cubic Spline Interpolation.

Step 3: Compute the mean of obtained upper envelope i.e., $S_{up}(t)$ and lower envelope i.e, $S_{low}(t)$.

$$N(t) = \frac{S_{up}(t) - S_{low}(t)}{2}$$

Step 4: The detail $D(t)$ of the given signal is obtained by taking the difference between the original signal $S(t)$ and the mean of the envelopes $N(t)$. $D(t) = S(t) - N(t)$

Step 5: Check whether the detail $D(t)$ is an IMF or not. ***Condition 1:*** If $D(t)$ have the same number of extrema and zero-crossings or can differ at most by one; ***Condition 2:*** If the average of $S_{up}(t)$ and $S_{low}(t)$ is zero at any point.

Step 6: If the above two conditions are fulfilled, then $D(t)$ is an IMF. So, replace $S(t)$ by its residue $R(t) = S(t) - D(t)$. If both conditions are not fulfilled, then $D(t)$ is not an IMF and thus, replace $S(t)$ by $D(t)$.

Step 7: Repeat step 1 to 5 until it meets the termination criteria.

Stopping Criteria:- For determining a stopping criterion for the sifting process, a standard deviation (SD) is used. The SD is calculated from the two successive sifting results and is generally set between 0.2 and 0.3.

$$S.D = \sum_{t=1}^{T} \frac{|S_n(t) - S_{n+1}(t)|^2}{S_n^2(t)}$$

D. Zhifang Wang et al. [2] face and iris recognition system

The fourth method taken for experimentation was proposed by Zhifang Wang *et al.* [2] who combined gabor and PCA features for recognition system. The detailed block diagram of the technique is given in figure 3. In the first phase, the features of iris and face are extracted and then, normalization procedure was carried out for applying to the fusion process.

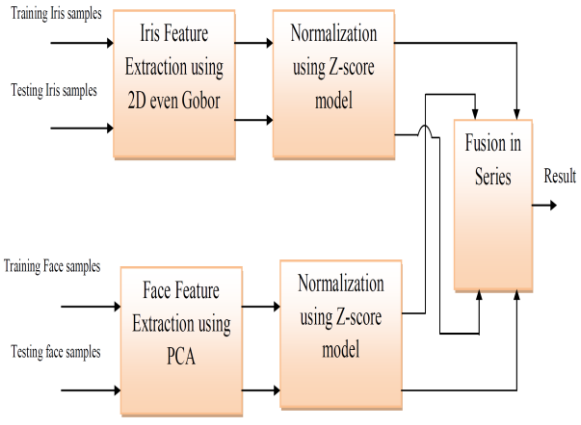


Fig. 3. Block diagram of the method proposed by Zhifang Wang *et al.* [2]

i) Face feature extraction: The following steps briefly portray the feature extraction method,

- Convert the matrixes of the face images into vectors as the training samples.

- Calculate the mean of all training sample images, $\bar{x}$.
- Create total scatter matrix S_T ,

$$S_T = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_j^i - \bar{x})(x_j^i - \bar{x})^T$$

Where, $\bar{x} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}x_j^i$ is the mean of all training sample images.

- Calculate the eigen values and the corresponding eigen vectors of S_T , and all eigen vectors compose the projection matrix W_{PCA} .
- Compute the projected feature vectors y_j^i ,

$$y_j^i = W_{PCA}^T x_j^i$$

- Determine the Euclidean distance between the projected feature vectors and the template vectors, and perform the classification.

ii) Iris feature extraction: For iris recognition, a Gabor filter is used, which is one of the prominent feature extractors. We segment the portion of the iris after performing the preprocessing process. Then, the features from the iris are extracted via 2D even Gabor filter. Gabor function contains fine characteristic in direction and frequency. A self-similar multi-channel filter family is obtained through rotation, scaling, and translation of the Gabor function. The 2D even Gabor filter is defined below,

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right]\bullet\cos(2\pi f\sqrt{x^2+y^2})$$

Where, σ_x, σ_y are the parameters of scale, and f represents the frequency. We partition the segmented iris image into sub-image with the same size, and correspondingly filter them by means of the Gabor filter family. The j^{th} filtered sub-image is achieved by

$$F_j(m,n)\iint I(x,y)g_i(m-x,n-y)dxdy$$

Where, $g_i(x,y)$ is the i^{th} even Gabor filter. Eventually, the absolute mean deviation is defined as the iris feature of the j^{th} filtered sub-image.

$$V_j = \frac{1}{N}\left(\sum_N F_j(m,n) - M\right)$$

Here, *M* is the mean value of the filtered sub-image, and *N* is the number of the filtered sub-image. All the absolute mean deviation of the filtered sub-image composed the iris feature. However, the dimension of the iris feature derived by Tan's technique is very large. Hence, a PCA is employed to overcome this problem and control the dimension of the iris feature equal to that of the face feature.

iii) Normalization**:** Normally, the feature-level fusion techniques directly combine two kinds of features after feature extraction process. It is well-known that the order of magnitude and the distribution between iris feature and face feature might be different due to the dissimilarity of the modal and extraction technique. In order to eradicate the unbalance and to obtain better performance, before fusion, the features are normalized by means of z-score model. Let a_j^i be a *d* -dimension iris feature of the j^{th} iris training sample from the i^{th} class, and b_j^i be a *d* -dimension face feature of the j^{th} face training sample from the i^{th} class. Subsequently, the iris feature set and the face feature set are represented as,

$$A = (a_1^1,...,a_m^1,a_1^2,...,a_m^n) \quad and \quad B = (b_1^1,...,b_m^1,b_1^2,...,b_m^n)$$

Let A_k be the k^{th} row of the iris feature set A. We use the following method to obtain the corresponding normalized component X_k . Initially, compute

$$C_k = \frac{A_k - \bar{A}_k}{\sigma_k}$$

Where, $\bar{A}_k$ represents the mean value of A_k , and σ_k is the standard deviation of A_k . Then, we can obtain the normalized component by

$$X_k = \frac{C_k - C_{\min}}{C_{\max}}$$

Where, $C_{\min}$ and $C_{\min}$ are the minimum and maximum values of C_k respectively. Finally, the normalized iris feature set is $X = (X_1,...,X_d)$. Similarly for face feature, repeat the same process and acquire the normalized feature set $Y = (Y_1,...,Y_d).$

iv) Feature Fusion: As we know, $X = (X_1,...,X_d)$ is a normalized iris feature vector, and $Y = (Y_1,...,Y_d)$ is a normalized face feature vector. The fusion feature ξ in sum rule can be defined as $\xi = (x_1 + y_1,...,x_d + y_d)$. For weighted sum rule, we take $\theta = \frac{3}{7}$ as the weighted parameter because the performance of iris recognition is superior compared to face recognition. The fusion feature in weighted sum rule is represented as $\xi = (x_1 + \theta y_1,...,x_d + \theta y_d)$. Thus, the sum rule can also be considered as a special case of weighted sum rule. In this paper, a series fusion technique is utilized. The format of the fusion feature is defined as $\xi = (x_1...x_d, y_1,...,y_d)$. This technique fuses two types of

feature into a long vector. Subsequently, a Euclidean distance is selected to categorize the fusion features.

III. Results And Discussion

The experimentation has been carried out on a 3.20 GHz i5 PC machine with 8 GB main memory running on a 64-bit version of Windows 2007. In this section, we analyze the performance of different technique. In sub-section 4.1, describes dataset description for face and iris recognition. Sub-section 4.2 describes the evaluation metrics using for our comparison work. The performance analysis results presented in sub-section 4.3 showed that the comparison results. The four techniques taken for comparative analysis are implemented in MATLAB.

A. Dataset Description

We have tested our approaches using two different datasets, namely CASIA and ORL. CASIA iris image database [3] includes 756 iris images from 108 eyes (hence 108 classes). For each eye, 7 images are captured in two sessions, where three samples are collected in the first session and four in the second session**.** In ORL database, there are ten different images of each of 40 distinct subjects. Face images and the Iris images from the databases are shown in figure 4 and 5 respectively.

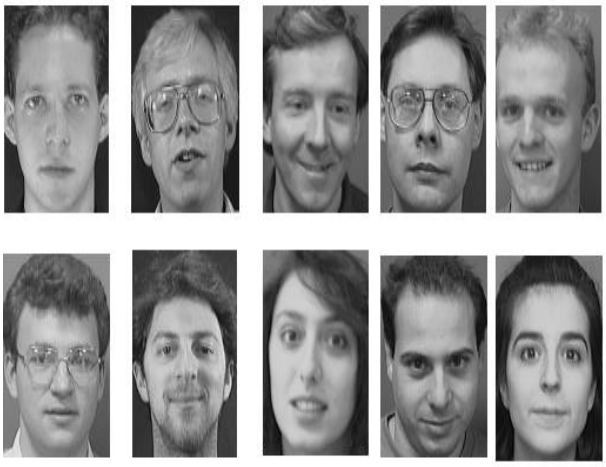


Fig. 4. Face images taken from ORL face database

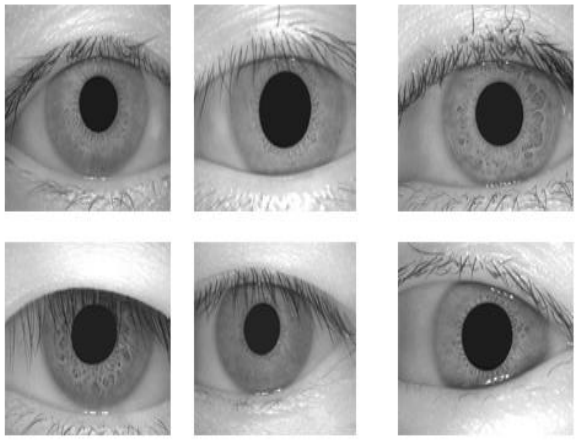


Fig. 5. Iris images taken from CASIA iris database

B. Evaluation Metrics

False Match Rate (FMR):- The False Match Rate (FMR) shows the proportion of persons who were falsely accepted during the characteristics comparison. Those efforts that were previously refused (Failure to Acquire, FTA) due to a low quality (e.g. of the image), contrary to FAR (False acceptance rate), are not taken into consideration. It depends on the application whether a falsely accepted feature contributes to increase the FAR or FRR (False rejection rate).

False Non-Match Rate (FNMR):- The False Non-Match Rate (FNMR) shows the proportion of persons who were falsely not accepted during the characteristics comparison. Those efforts that were previously rejected (Failure to Acquire, FTA) due to a low quality (e.g. of the image), contrary to FRR, are not taken into consideration. Again, it relies on the application whether a falsely non-accepted feature contributes to increase the FRR or FAR.

ROC curve: The $FMR(t)$ (False Match Rate) and $FNMR(t)$ (False Non-Match Rate) are calculated for t ranging from 0 to 1. Then, the ROC curve is plotted FMR vs. FNMR for varying threshold t. The plotted ROC curve is extensively used in the contest to compare the performance of different algorithms. One more parameter used for comparison is, Equal Error Rate (EER) that is computed as the point where $FNMR(t) = FMR(t)$.

C. Performance Analysis

1) Analysis of the LGXP-based Approach for Features Fusion in Face and Iris recognition

The performance analysis of the LGXP is presented in this sub-section. For various threshold values, FMR and FNMR are computed. Then, the graph is plotted for the computed values to find the efficiency in acceptance of the genuine user and the rejection of the impostor user for different threshold levels. The performance analysis graph on LGXP method is shown in figure 6. From the plotted graphs in figure 6, we have found the equal error rate (EER) of the LGXP- approach is given as, EER= 0.5.

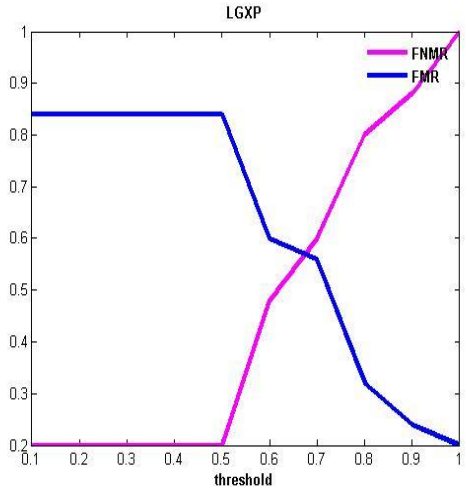


Fig. 6. Performance analysis graph for LGXP

2) Analysis of the LBP-based Approach for Features Fusion in Face and Iris recognition

The performance analysis of the LGXP-based approach is presented in this sub-section. For various threshold values, FMR and FNMR are computed. Then, the graph is plotted for the computed values, shown in figure 7. From the plotted graphs in figure 7, we have found the equal error rate (EER) of the LBP-based approach is given as, EER= 0.5.

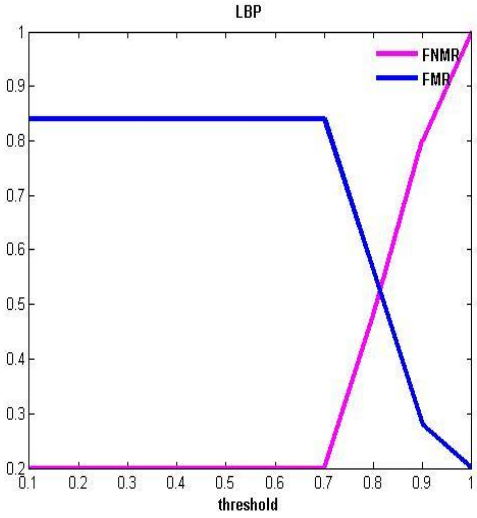


Fig. 7. Performance analysis graph for LBP

3) Analysis of the EMD-based Approach for Features Fusion in Face and Iris recognition

This section presents the performance plot of the EMD-based approach in face and iris recognition. For various thresholds, the FMR and FNMR is computed for the same approach and the computed values are used to plot graph. The graph shown in figure 8 is the performance of the EMD-based approach such a way, the EER of the technique is computed from the plotted graph, EER= 0.56.

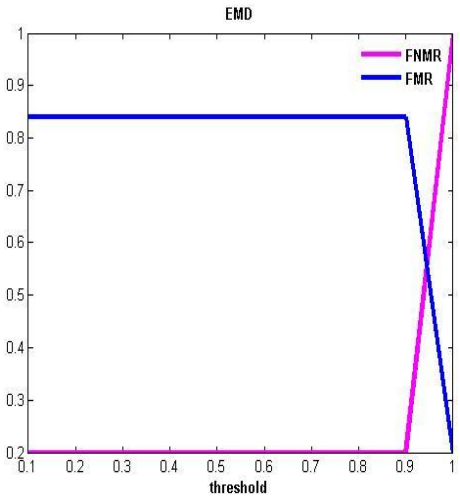


Fig. 8 Performance analysis graph for EMD

4) Analysis of the PCA-based Approach for Features Fusion in Face and Iris recognition

The performance analysis graph of PCA-based technique is shown in figure 9. From the plotted graphs shown in figure 9, we have found the equal error rate (EER) of the approach, EER= 0.57 which is high compared with other three techniques.

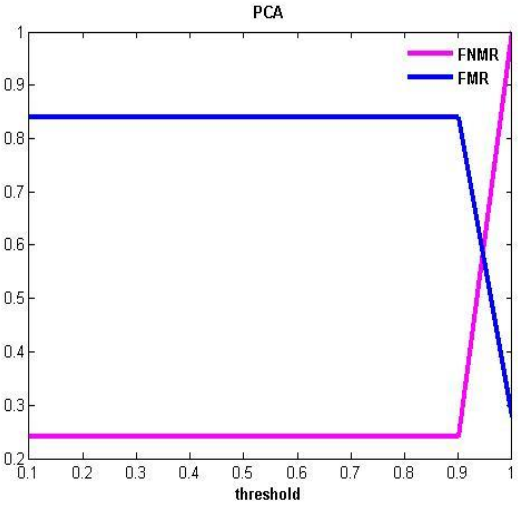


Fig. 9. Performance analysis graph for PCA

D. Comparative Analysis

The comparative analysis of four multimodal approaches such as LGXP, LBP, EMD and PCA is performed using the ROC curve. The plotted ROC curve for the comparative analysis is given in the figure 10. From the ROC curve graph, we have analyzed that the LGXP and LBP-based method has a lower FNMR value which means the better security of the proposed multimodal biometric system. When compared with the PCA method, the proposed three approaches provided better results.

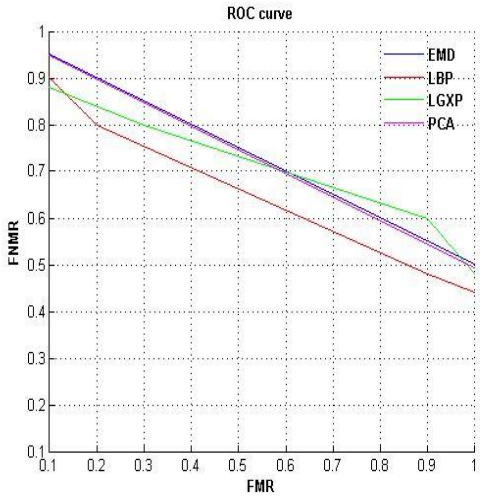


Fig. 10 ROC curve graph for four methods

IV. CONCLUSION

We have presented comparative analysis of four different feature extraction approaches, such as LBP, LGXP, EMD and PCA. The major processes of the four methods are consists of 1) Feature extraction from face image, 2) Feature extraction from iris image and 3) Fusion of face and iris features. The performance of the feature extraction methods in multi-modal recognition was analyzed using FMR and FNMR to study the recognition behavior of these approaches.

Then, an extensive analysis was carried out to find the effectiveness of different approaches using ROC curve. For experimentation, we have used the *CASIA* iris image database and the ORL face database and the evaluation results clearly demonstrated that the performance of the LBP and LGXP method is better compared with PCA-based technique.

REFERENCES

[1] Deepak Sharma and Dr. Ashok Kumar "Multimodal Biometric Recognition System: Fusion of Face and Iris Features Using Local Gabor Patterns**",** International Journal of Advanced Research in Computer Science (IJARCS), Volume 2, Number 6, Nov-Dec 2011.

[2] Zhifang Wang, Erfu Wang, Shuangshuang Wang and Qun Ding, "Multimodal Biometric System Using Face-Iris Fusion Feature", journal of computers, vol. 6, no. 5, May 2011.

[3] Chinese Academy of Sciences - Institute of Automation (CASIA), Database of 756 Gray scale Eye Images, http://www.cbsr.ia.ac.cn/IrisDatabase.htm,Version 1.0, 2005.

[4] Hayasaka, A.Shibahara, T.Ito, K.Aoki, T. Nakajima and H. Kobayashi, "A 3D Face Recognition System Using Passive Stereo Vision and Its Performance Evaluation", In proceedings of International Symposium on Intelligent Signal Processing and Communications, pp.379 – 382, 2006.

[5] Ching-Han Chen and Chia TeChu, "Fusion of Face and Iris Feature for Multimodal Biometrics", Lecture Notes in Computer Science, Vol. 3832, pp. 571-580, 2005.

[6] Sanjay Kumar Mohanty and Prasant Kumar Pattnaik, "Authentication Based on Texture Analysis and SVM Classification", International Journal of Instrumentation, Control & Automation (IJICA), Vol.1, No.1, pp. 61-66, 2011.

[7] A. Ross and A. K. Jain, "Information fusion in biometrics, "Pattern Recognition Letters, Vol. 24, pp. 2115–2125, Sep 2003.

[8] L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin, "Is independence good for combining classifiers?," in Proc. of Int'l Conf. on Pattern Recognition(ICPR), vol. 2, pp. 168–171, 2000.

[9] Arun Ross and Anil K. Jain, "Multimodal Biometrics: An Overview", in Proc. of 12th European Signal Processing Conference (EUSIPCO), pp.1221-1224, September 2004.

[10] Pascal Paysan, ReinhardKnothe, Brian Amberg, Sami Romdhani and Thomas Vetter, "Face Recognition Using 3-D Models: Pose and Illumination ", In proceedings of IEEE , Vol. 94, No.11, pp.1977 - 1999, 2009.

[11] H B Kekre and V ABharadi, "Gabor Filter Based Feature Vector for Dynamic Signature Recognition", International Journal of Computer Applications, Vol. 2, No.3, pp.74-80, May 2010.

[12] A. Jain, L Hong and S Pankati. "Biometric Identification". Communications of The ACM, vol. 43, no. 2, pp. 90-98. Feb. 2000.

[13] J.G.Daugman, "High Confidence Visual Recognition of Persons by a Test of Statistical Independence", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.15, No.11, pp. 1148–1161, 1993.

[14] Sandipan P. Narote, Abhilasha S. Narote and Laxman M. Waghmare, "Iris Based Recognition System using Wavelet Transform", International Journal of Computer Science and Network Security, Vol.9, No.11, November 2009.

[15] Sudha Gupta, Viral Doshi, Abhinav Jain and SreeramIyer ,"Iris Recognition System using Biometric Template Matching Technology", International Journal of Computer Applications , Vol.1, No.2, pp.21-30, 2010.

[16] Srinivasa Kumar devireddy, "An Accurate Human Identification through Iris recognition", Georgian electronic scientific Journal in Computer Science and Telecommunication, Vol.6, No. 23, pp. 22-29, 2009.

[17] Prof G.Ramaswamy VudaSreenivasarao, Dr.P.Ramesh, D.RaviKiran, "A Novel Approach for Human Identification through Fingerprints", International Journal of Computer Applications, Vol.4, No.3, pp. 35-42, July 2010.

[18] Z.Yaghoubi, K.Faez, M.Eliasi and A.Eliasi, "Multimodal biometric recognition inspired by visual cortex and Support vector machine classifier", International Conference onMultimedia Computing and Information Technology (MCIT), pp.93 – 96, 2010.

[19] Anil Jain, Karthik Nandakumar and Arun Ross, "Score normalization in multimodal biometric systems", Pattern Recognition, Vol. 38, No.12, pp.2270-2285, Dec. 2005.

[20] L. Hong, A.K. Jain, S. Pankanti, " Can multibiometrics improve performance? ", In proceedings of IEEE Workshop on Automatic Identification Advanced Technologies,pp.59-64,1999.

[21] Robert Snelick, Umut Uludag, Alan Mink, Michael Indovina, and Anil Jain, Large-Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 3, pp.250-255, March 2005.

[22] A.K. Jain, R. Bolle, and. Pankanti, "Biometrics: Personal Identification in Networked Society", 1999.

[23] ShufuXie, ShiguangShan, XilinChenand Jie Chen, "Fusing Local Patterns of Gabor Magnitude and Phase for Face Recognition", IEEE Transactions on Image Processing, Vol. 19, No. 5, pp.1349-1362 , May 2010.

[24] JieLin, Jian-Ping Li, Hui Lin and Ji Ming," Robust person identification with face and iris by modified PUM method", International Conference on Apperceiving Computing and Intelligence Analysis, pp. 321– 324, 2009.

[25] Jun-Ying GanandJun-Feng Liu, "Fusion and recognition of face and iris feature based on wavelet feature and KFDA", International Conference on Wavelet Analysis and Pattern Recognition, pp.47-51, 2009.

[26] Shoa'aJadAllah Al-Hijaili and ManalAbdul Aziz, "Biometrics In Health Care Security System, Iris-Face Fusion System", International Journal of Academic Research, Vol. 3, No. 1., pp.11-19, January, 2011.

[27] Baochang Zhang, Shiguang Shan, XilinChenand Wen Gao, "Histogram of Gabor Phase Patterns (HGPP): A Novel Object Representation Approach for Face Recognition", IEEE Transactions on Image Processing, Vol. 16, No. 1, pp.57-68, January 2007.

[28] HarinSellahewa and Sabah A. Jassim, "Image-Quality-Based Adaptive Face Recognition", IEEE Transactions on Instrumentation and Measurement, Vol. 59, no. 4, pp.805-813, April 2010.

[29] Zhenan Sunand Tieniu Tan, "Ordinal Measures for Iris Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.31, No.12, pp. 2211 – 2226, 2009

[30] Ryan N. Rakvic, Bradley J. Ulis, Randy P. Broussard, Robert W. Ives and Neil Steiner, "Parallelizing Iris Recognition", IEEE Transactions on Information Forensics and Security, Vol. 4, No. 4, pp. 812- 823, December 2009.

[31] Ojala, T., Pietik¨ainen, M., Harwood, "D.: A comparative study of texture measures with classification based on feature distributions," Pattern Recognition, vol. 29, pp.51–59, 1996.

[32] D. J. Duffy, "The Application Of Hilbert-Huang Transforms To Meteorological Datasets", Journal of Atmospheric and Oceanic Technology, vol. 21, no. 4, pp.599–611, 2004.

An Approach to Keep Credentials Secured in Grid Computing Environment for the Safety of Vital Computing Resources

Avijit Bhowmick
Asst. Professor, Department of CSE/IT
Dr.B.C.Roy Engineering College
Durgapur, West Bengal, India

C T Bhunia
Director
National Instituite of Technology
Yupia, Arunachal Pradesh, India

***Abstract*—Presently security attacks have aimed to vulnerabilities in repetitive-use authentication secrets like static passwords. The passwords are used by user in clients side are vulnerable, as the attackers can gain access to a user's password using different types of viruses as it is being typed. These attacks are directing many Grid sites to explore one-time password solutions for authentication in Grid deployment. We present here a novel mechanism called N-LSB where Grid security will be integrated with modified LSB based steganographic technique in order to meet the higher security demands for Grid credentials.**

Keywords- Grid; security; LSB; authentication.

I. INTRODUCTION

Huge network bandwidth, more potent software and hardware of computer at low cost and the flooding of Internet has dived towards need of the latest powerful computing environment. In the late 1990's, Ian Foster et.al. Proposed a complex robust computing environment named Grid computing which was a combination of software and hardware infrastructure where resources are shared for the purpose of computing and storage [1, 2].Because of huge initial investment and high maintenance cost of mainframe or supercomputer, most of the organizations are unable to utilize these type of technologies and again internal resources of the organization also cannot be utilized completely[3].Grid computing technology is defined as "A Computational Grid is a hardware and software infrastructure that provides dependable, consistent, pervasive and inexpensive access to high and computational capabilities"[4].It refers to the combination of dispersed computer resources from several administrative domains working together to reach a common goal what distinguishes it from conventional high performance systems such as cluster computing [5].

High performance heterogeneous resources like processors, network etc. of different domains are utilized for different Grid applications through Grid middleware. The Globus[12]. Gridbus[15] etc. are the examples of Grid middleware.

Grid computing resources differ from modest number of clusters like the TeraGrid [6] to millions interconnected Pcs like SETI@Home [7].

Since, resources developing an inclusive set of mechanisms and policies for safeguarding the Grid is most essential challenge for Grid computing dubiously in some ways. At present, Grid security analysis and development turns around developing higher solutions to require care of the subsequent requirements: Authentication, Secure data Communication, and effective Security Policies, Authorization, and resource access management. Resources and data security are two basic needs in Grid applications [8, 9]. Coordinated resource sharing and downside breakdown in dynamic, multi-domain virtual organizations are the actual and exact issues that lie behind the Grid conception [10].Intruders attack Grid applications through malicious code, may lead to ruin all the running applications at the same platform. The concept of coordinated resource sharing behind Grid concept is just not only simple file exchange, but also direct access to the vital resources of Grid [11]. The resource administration in Grid environment is difficult due to: (a) geographical dispersion of resources,(b)resource heterogeneousness, (c)having different resource island with their own policies(d) Grid domains inequity [14].

Currently, security is built into Grid computing toolkits like Globus toolkit [12] which is employed at the resource provider sites. The toolkit manages secure channels, authentication [13], resource login, delegation, and resource handling [11]. Among all other issues authentication is the first and most important process of Grid Security Infrastructure (GSI).(GSI) is based on asymmetric cryptography used in a "Public Key Infrastructure" (PKI). Asymmetric cryptography allows users to communicate securely without the necessitate for a previous secret channel to exchange encryption key. Exploiting features of a specific class of mathematical challenges that are simple to generate but virtually impossible to solve (like factorizing large prime numbers), end-entities generate a Complementary set of keys: a "private key" that will be kept secret and a "public key" that is broadcast to the world. Data encrypted with the public key can only be deciphered with the private key (and vice versa). Thus data confidentiality, message integrity and non-repudiation can be attained between two halves of the key pair.

There are many apprehensions about the problems of the current user-managed identity credentials in the Grid PKI. It is essential that the user's private key, encrypted on disk, is

correctly protected both in terms of file protections and in the excellence of the pass-phrase used to encrypt it. As is often the case in the management of user-held Ssh private keys, this does not always happen. The fact that the management of certificates and private keys inside web browsers is also very complex that does not develop the situation.

For this above discussed issue we are going to implement a novel technique developed by us called N-LSB (Nearest Least significant Bit) which is LSB based steganographic advanced modified substitution technique for enhancing security in Grid.

II. LSB Substitution Technique Overview

Substitution based steganography replaces redundant or wasted bits of a digital cover file (digital image, audio or video file) with the bits from the secret message. Let three pixels of RGB scheme as follow:

(00100111 11101001 11001000)
(00100111 11001000 11101001)
(11001000 00100111 11101001)

Changing the first bit i.e. MSB from a 0 (00100111) to a 1 (10100111) will radically change the color; anybody can detect the change with the help of naked eyes. But changing the LSB from 1 (00100111) to a 0 (00100110) cannot be distinguishable in open eyes.

Take a look at how LSB substitution method is applied in steganography to hide a message.

When the character "A" (10000011) is embedded the changed Pixels are

(00100111 11101000 11001000)
(00100110 11001000 11101000)
(11001001 00100111 11101001)

Here for the eight byte information only four have been changed. In best case not a single byte will be changed. But in worst case for eight byte information value of the three pixels will be changed. Sometimes, depending on the pixel, adjusting the LSB can dramatically affect the pixel's properties, making it look out of place in the picture, and therefore subject to detection. This problem can limit the amount of substituted bits, and therefore the size of the secret message. From the above discussion it is clear that keeping the image's properties intact the image size must be eight times greater than the secret message. This is one of the drawbacks of LSB technique. Our N-LSB (Secured-LSB) technique can overcome this problem.

III. Proposed And Developed N-Lsb Method For Keeping Secured In Grid

This new approach not only overcomes the above-mentioned problem but it has an extra layer of security. Like LSB technique N-LSB also require one text file for storing the secret text message and one multimedia carrier file (such as image file, an audio or a video file).

After embedding the source text file information inside each 'atom' of the carrier multimedia cover file, a 'Stego File' is generated whose file type is same as the source cover file. This stego file is used to decrypt the hidden information. N-LSB technique mainly subdivided into two parts: first part, normally done at the sender side is for Encryption of the Source Text File inside the Cover file. The second part, normally done at the receiver side is for Decryption of the secret information from the Stego File.

A. Encryption Method of N-LSB technique:

Step-I

Authenticity checking of the Computer to protect unauthorized use of the source program. This authentication of the system ensures that there will be no misuse of code.

Step-II

Initialization vector is set to the password and is packed into an array.

Step-III

Initialization vector is XOR-ed with the current plain text block to generate the cipher text block & the cipher text stored in a temporary file with a unique file name in the current directory without any kind of user interaction.

Step-IV

This temporary file is actually used to embed the cipher text (stored in it) in a special manner that is based on the old LSB method; but in N-LSB instead of using LSB as the index to embed the text message we generate the index by using Double Hashing Technique.

Step-V

If the source cover file is much smaller than the size of the text information file apply Roll-Back Mechanism to avoid unsuccessful encryption as well as zero probability decryption even if both encryption and decryption password are same.

Here we introduce the variable key generation for the Double Hashing and the value of the key varies depending on the nature of the media files; this does not mean that the key value remains the same for a particular type of media file. The value of the key is different for different files. The 'N-LSB' method actually generates the corresponding indexes that are nearer to the LSB for embedding the text messages. The index may be sometime (LSB-2), sometimes (LSB-1) or sometime LSB itself. If the generated index becomes LSB-2 then LSB-2, LSB-1 and LSB are used, if index becomes LSB-1 then LSB-1 and LSB positions are used, and if the generated index is LSB then only the LSB position is used to embed the secret text information. Using this special feature we try to overcome the above mentioned 1:8 ratio problem.

B. Decryption Method of N-LSB Technique:

Step-I

Authenticity checking of the Computer to protect unauthorized use of the source program. This authentication of the system ensures that there will be no misuse of code.

Step-II

Providing only the correct password (same as encryption password) proper decryption is possible.

Step-III

Using the correct password, the encrypted cipher text is again converted into plain original text in- formation by applying the reverse method already discussed briefly above used to encrypt the secret in- formation. Another exclusive temporary file is used to hold the intermediate results. This temporary file is used to generate the secret information that was sent by the authentic sender.

Step-IV

The two temporary files are deleted as those are no longer needed. We call it Level Zero Destruction.

B. Explanation of the proposed N-LSB Technique :

For each bit of each CIPHER character of the UNIQUE TEMPORARY files, first four bits of the pixel

1. MSB
2. (MSB+1)
3. (MSB+2)
4. (MSB+3)

are added to generate the HASH-KEY hk (for each bit of each CIPHER character of the Unique Temporary file). Clearly (LSB-3) is not involved in calculation in our N-LSB method, we call it a Barrier. It acts as a Barrier between the calculating (MSB, MSB+1, MSB+2, MSB+3) and the calculated (LSB, LSB-1, LSB-2) bits.

Next we follow the following Double Hash Function on each pixel value (for each bit of each CIPHER character of the Unique Temporary file) -

```
int hh=6;
int hash index  hi = (hask key hk % hh ) + 1;
int hm  = (hash key hk % (hh - 2) ) + 1;
int index  = abs( ( (hash index hi + hm) % hh) + 1 );
float ii = (float)index;
```

Finally we check the Ultimate Index (int index) is greater than 5 (Barrier value i.e LSB-3) or not to ensure the Min value of the Ultimate Index (int index) must be 6. If not multiply the hashed index (float ii) with a fractional value 1.55 (i.e. float hmf) and then convert the hashed index (float ii) it to its nearest integer value and stores it to the Ultimate Index (int index).

The above step is continued until Ultimate Index (int index) is greater than the Barrier Value (i.e. 5, i.e. LSB-3).

Another checking is there to ensure that the max value of the Ultimate index (int index) is 8 (i.e. the LSB of that pixel).

One thing is to notice that Ultimate Index (int index) is taken as an Integer while the Hashed Index (float ii) is taken as Float. This is done to calculate the Ultimate Index (int index) more significantly and mathematically eliminating the 0.5 ERROR.

For example if the Hashed Index ii =6.15 the Ultimate Index (index) will be 6 and not 7 but if the Hashed Index ii =6.65 the Ultimate Index (index) will be 7 and not 6.This is called 0.5 Error Elimination. Clearly we have to do the above calculation on each pixel of the Source Cover file for each embedding that is we have to do the above calculation for each bit of each CIPHER character of the Unique Temporary file.

In an Authenticated Machine after checking the Length Error (LE) create a UNIQUE TEMPORARY (UT) file that will contain the Cipher text. UNIQUE in the sense that there should be no name-space collision with any other pre-existing files in the currently working directory and this is assured by the program code. The UNIQUE TEMPORARY file can have any extension. Length Error (LE) occurs if the length of the Text file is larger than the Source Cover file.

We have introduced here a brand-new concept ENCRYPTION PASSWORD where each character of the original text is repeatedly XOR-ed with the characters of the given password to generate the CIPHER text.

For example: one text file contains the data: 1 2 3 4 and user's password is: ab

First 1 is XOR-ed with a

Then 2 is XOR-ed with b

Then 3 is again XOR-ed with a

Ultimately 4 is XOR-ed with b.

From now there is no need of the original text file because now we have got the unique temporary file.

This is done :

a) To remove the dependency factor (of the program on the original source text file).

b) To increase the security level (because we are encrypting the CIPHER text not the ORIGINAL text).

Introducing new concept of password based security where one must know the ENCRYPTION password to decrypt the original text. Until and unless one knows the password given at the time of encryption he is unable to decrypt the original text. Actually, there is no such concept of checking the authenticity of the given password as the old Password concept. If wrong password is entered then decryption will be wrong and the output text is wrong without showing you any ERROR.

Again,in our N-LSB (Nearest List Significant Bit) Method, Double Hashing Technique is used to generate the INDEX (EMBEDDING Location) for each bit of each CIPHER character of the UNIQUE TEMPORARY file. Index can be LSB, (LSB-1) or (LSB-2) of the pixel value.

We follow one of the following rules (i.e., if it satisfies one rule the next rule is not checked):

a. If the Ultimate Index is 6 (i.e., LSB-2) then we replace the 6^{th} (i.e., LSB-2), 7th (i.e., LSB-1) and 8^{th} (i.e, LSB) bit position of the same pixel with the three consequent bits of the Cipher character.

b. If the Ultimate Index is 7 (i.e, LSB-1) then we replace the 7^{th} (i.e., LSB-1) and 8^{th} (i.e., LSB) bit position of the same pixel with the remaining two consequent bits of the Cipher character.

c. If the Ultimate Index is 8 (i.e., LSB itself) then we replace only the 8th (i.e., LSB) bit position of the pixel with only one bit of the Cipher character.

Finally embed each bit of each CIPHER character of the Unique Temporary file in the Ultimate Index or the Double Hashed Index of each pixel of the Source Cover file to generate the Destination File. After the successful Encryption the Unique Temporary (UT) is deleted because it no longer needed. We call it Level Zero Destruction**.**

IV. Performance Analysis And Discussions

Even if Stego file is detected by attacker then also he cannot be able to decrypt the Stego file due to authentication failure because it does not match the computer's environmental specification which was combined with password which adds an extra layer of security. Unfortunately if he gets the correct password then he must have to know the encryption algorithm of N-LSB method which is more secured than so called LSB substitution method which enhances the level of security. If the algorithm is found out one must know the process of variable key generation algorithm using double hashing technique. It is quite difficult to find out the hash functions that have been used.

LSB Technique	Proposed N-LSB Technique
Exists 1:8 ratio problems.	Removed 1:8 ratio problems.
Only LSBs are used, so less information can be embedded.	Nearest bits of LSBs are also utilized, so in same cover file more information embedded.
Plain text information is directly kept in LSBs, less secured.	Plain text encrypted first, then kept in nearest LSBs, complete security for data.
For maximum information embedding in cover file, all the LSBs are utilized.	For the same amount of information embedding in same cover file 37.5 % less LSBs are needed.

Fig.1 Performance comparison

Again, in LSB substitution technique to embed a character eight bytes must be extracted from the cover file. But in N-LSB method eight bytes needed in worst case if and only if the generated index is only LSB in all cases. In best cases, if it produce index LSB-2 then only three bytes will be extracted from the cover file. So in best case, it requires only 37.5% of the cover file compared so we can increase the embedding capacity more than 60%. Obviously in average case, N-LSB method requires 50% less size of cover file and it increase 50% embedding capacity.

We can also increase the embedding capacity by converting the entire secret message into either lower case or in uppercase. With these multilevel of security the proposed N-LSB method is secured from attacks than any other existing techniques.

V. Conclusion

Vital information through LSB technique reduces the probability of detection of the information unless knowing the cryptographic algorithm behind the mechanism. Utilizing this unique feature, we have implemented through our developed unique method called N-LSB Technique where we have observed less time and less no of bits are needed for keeping users credentials secured. So, this robust technique is very much useful to keep Grid credentials secured for ensuring data integrity and information security in Grid computing environment for issue of accessing vital heterogeneous resources in secured manner.

References

[1] F. Berman, G. Fox and T. Hey (eds.), Grid Computing: Making the Global Infrastructure a Reality. Wiley, 2003.

[2] M. Cosnard and A. Merzky, "Meta- and Grid-Computing", in Proceedings of the 8th International Euro-Par Conference, August 2002, pp. 861–862.

[3] Nadia Ranaldo, Eugenio Zimeo. A Framework for QoS- based Resource Brokering in Grid Computing. In 5th IEEE ECOWS, the 2nd Workshop on Emerging Web Services Technology, Halle, Germany, 2007.

[4] I. Foster, C. Kesselman, and S. Tuecke, "The anatomy of the Grid: Enabling scalable virtual organizations." Int. J. Supercomputing, vol. 15, no. 3, pp. 200-222, 2001.

[5] Foster, I. and Kesselman, C. Computational Grids. Foster, I. and Kesselman, C. eds. The Grid: Blueprint for a New Computing Infrastructure, Morgan Kaufmann, 1999, 2-48.

[6] "National Science Foundation TeraGrid". From http://www.teraGrid.org.

[7] SETI@Home: The Search for Extraterrestrial Intelligence. http://setiathome.ssl. berkeley.edu/

[8] F. Berman, R. Wolski, H. Casanova, W. Cirne, H. Dail, M. Faerman, S. Figueira, J. Hayes, G. Obertelli, J. Schopf, G. Shao, S. Smallen, N. Spring, A. Su and D. Zagorodnov, "Adaptive Computing on the Grid Using AppLeS", IEEE Trans. on Parallel and Distributed Systems, Vol. 14, April 2003.

[9] V.Welch, F. Siebenlist, I. Foster, J. Bresnahan, K. Czajkowski, J. Gawor, C. Kesselman, S.Meder, L. Pearlman and S. Tuecke, "Security for Grid Services", in Proceedings of the HPDC-12, 2003.

[10] Zhiguo Shi, Yeping He, Xiaoyong Huai, Hong Zhang. Identity Anonymity for Grid Computing Coordination based on Trusted Computing. Proceedings of the Sixth International Conference on Grid and Cooperative Computing. pp. 403-410, 2007.

[11] Foster, I., Kesselman, C., Tsudik, G. and Tuecke, S. A Security Architecture for Computational Grids. ACM Conference on Computers and Security, 1998, pp: 83-91.

[12] I. Foster. Globus toolkit version 4: Software for service- oriented systems. In Proc. of the IFIP International Conference on Network and Parallel Computing, 2005.

[13] J. Basney, W. Nejdl, D. Olmedilla, V. Welch, and M. Winslett. Negotiating trust on the Grid. In 2nd Workshop on Semantics in P2P and Grid Computing, New York, May 2004.

[14] Farag Azzedin, Muthucumaru Maheswaran, "Towards Trust-Aware Resource Management in Grid Computing Systems," ccGrid, p. 452, 2nd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID'02), 2002.

A Novel Architecture for Network Coded Electronic Health Record Storage System

B. Venkatalakshmi S. Shanmugavel

***Abstract*— The use of network coding for large scale content distribution improves download time. This is demonstrated in this work by the use of network coded Electronic Health Record Storage System (EHR-SS). A Novel Architecture of 4-layers to build the EHR-SS is designed. The application integrates the data captured for the patient from three modules namely administrative data, medical records of consultation and reports of medical tests. The lower layer is the data capturing layer using RFID reader. The data is captured in the lower level from different nodes. This data is combined with some linear coefficients using linear network coding. At the lower level the data from different tags are combined and stored and at the level 2 coding combines the data from multiple readers and a corresponding encoding vector is generated. This network coding is done at the server node through small mat lab net-cod interface software. While accessing the stored data, the user data has the data type represented in the form of decoding vector. For storing and retrieval the primary key is the patient id. The results obtained were observed with a reduction of download time of about 12% for our case study set up.**

Keywords- Network coding; Electronic Health Record system; Distributed content storage; RFID; Linear network code.

I. INTRODUCTION

The fundamental idea of Network Coding spreads its potential in various network performance metrics, in the last decade. The use of network coding for large scale content distribution to improve download time [1] is the motivation for this work of network coded Electronic Health Record (EHR) Storage System. Other initiatives of network coded content distributions are for, network coding based distributed storage in Sensor Networks [2], reliability improvement under server failures[3], and more like this.

Today's widespread use of social networks, logs, videos, mails etc., makes an exponential increase in data storage. Application specific data centers like supply chain process of an Industry, Health Care systems of Hospitals, Academic records of Educational Institutions etc., are also require maintenance of large volume of data. To obtain more reliable data storage for such cases, distributed data storage becomes essential. Distributed file storage in internet uses large data centers like Ocean store [4], Total Recall [5] etc. The entire system design requires offering reliable access to data.

Nowadays machine generated data exceeds human generated data and we need storage systems of the order of Exabyte's. When we consider distribution of data in wireless networks, there exists challenges like wireless data rates, link failures and packet loss probability. Thus we need to develop strategies by deployment and integration of newer technologies. The desirable performance metrics includes rebuild time, read/write bandwidth and storage efficiency. The major pulling factor is the tradeoff between reliability and redundancy. Intelligent architectures are required to achieve this and one such effort is this work by focusing on the metric rebuild time.

We make use of content based network coding for a selected scenario of EHR, which results in better storage and retrieval of contents. The rest of the paper is organized as follows. Section 2 reveals the related works and the basic principle used in our work. In section 3 we propose a new architecture for the design of network coded EHR-SS. Section 4 elaborates our case study set up and network coding strategy. The outputs obtained are described in section 5 and finally concluded in section 6.

II. RELATED WORKS

Network Coding allows more intelligence at the nodes to perform simple computation (encoding). The data packets are combined and stored for distributed storage. Also the profit of network coding is achieved using linear transformations. Various research works have illustrated the benefits of network coding and the design of network coding for wired and wireless networks. The challenge in distributed storage arises when both the data sources and source nodes are distributed. A survey on the usage of network coding for distributed storage in wireless sensor networks [6] paved a new way of research in the application of network coding. Network Coded Distribution storage systems with storage and repair bandwidth tradeoff is another milestone in the use of network coding.

File download time of almost 20-30% improvement is achieved by Christor Gkantsidis et al in [1] by the use of network coding for large scale content distribution. The results are tested in heterogeneous networks. To illustrate how network code improves the propagation of information without a global coordinated scheduler we consider the following (simple) example. In Figure 1 assume that Node A has received from the source packets 1 and 2. If network coding is not used, then, Node B can download either packet 1 or packet 2 from A with the same probability. At the same time that Node B downloads a packet from A, Node C independently downloads packet 1. If Node B decides to receive packet 1 from A, then both Nodes B and C will have the same packet 1 and, the link between them cannot be used.

If network coding is used, Node B will download a linear combination of packets 1 and 2 from A, which in turn can be used with Node C. Obviously, Node B could have

downloaded packet 2 from A and then use efficiently the link with C.

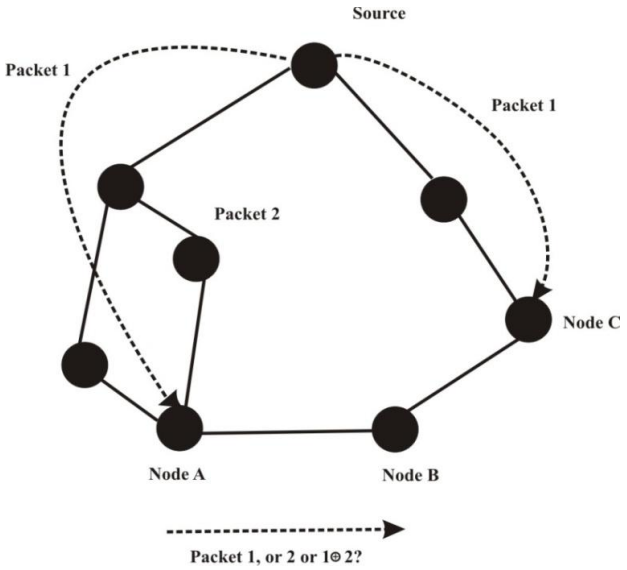


Figure 1. Network coding benefits

III. PROPOSED ARCHITECTURE FOR EHR-SS

Effective utilization of EHRs demands convergence of technologies, adoption of smart telephones and social media [7]. In [8] Ansell predicted the need for provincial licensing bodies for the tools of Social Networks. Followed by this, Health Care Networks evolved only as a source to share user's queries as a general forum. This capacity can be expanded for discovering the misuse of medical records as per the statements of You Chen [9]. Stanley Feld [10] insisted on the need to practice Evident Based Medicine (EBM) which is a feasible practice by the merge of Social Media with Health Information Systems. The unexplored potentials of Social media concepts in diabetes self-management on mobile devices are highlighted by [11].The above survey provide a thrust for our model of EHR-SS.

We propose an architecture of 4-layers to build the EHR-SS as shown in Figure 2. The lower layer is the data capturing layer. This uses RFID passive tags. The captured details are uploading on the client systems in the 2nd layer. The networked clients are connected in this layer and they upload the data to the server in the 3rd layer. Network coded details exists in both 2nd and 3rd layer. The data cloud is optional and is constructed in the 4th layer of the architecture.

A. RFID based Data Capture

The RFID system consists of a reader, tag and the host system. The reader and tag communicate through a RF signal link. Figure 3. shows components of a reader and the tag. The reader is the central nervous system of the entire RFID hardware system, establishing communication with the tag and the host system. The reader may be fixed (or table model) or handheld which is relatively costlier. The RFID tag is a device that can store and transmit data to a reader in a contactless manner using radio waves. RFID tags could be passive, active or semi-active (or semi passive).

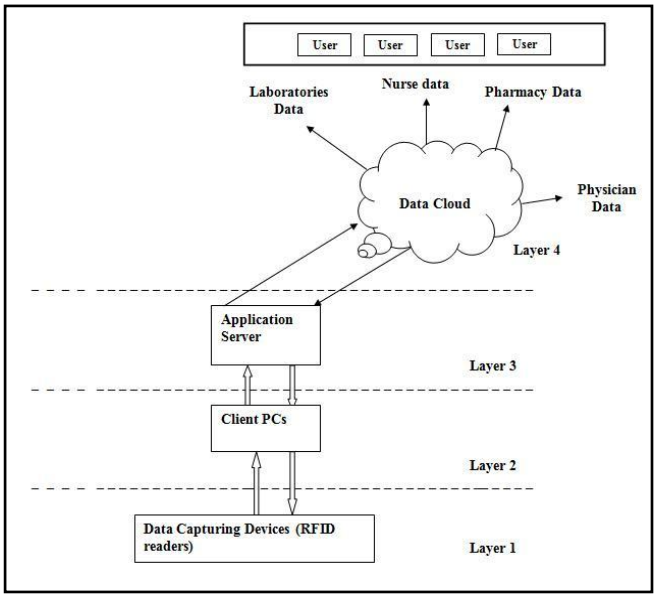


Figure 2. Architecture of EHR-SS

A passive tag has no on board power source and hence they have a limited range. They are also cheaper compared to an active or semi – active tag. The tag may also be a read only (RO), write once, read many (WORM) and Read – Write (RW) types. Both active and passive tags can be RO, or WORM or RW types. The RFID reader operates at the following frequencies, low frequency (125 KHz), high frequency (13.56 MHz), Ultra High Frequency (868 and 915MHz) and microwave frequency's (either 2.45 GHz or 5.8 Hz). A typical LF RFID system operates at 125 KHz or 134.2 KHz.

These RFID system have low data transfer rates from the tag to the reader and specially good if the operating environment contains metals liquids, dist, snow operating environment contains metals, liquids, dist, snow or mule. But the read range is low and need larger antennas refueling in higher cost tags further the tag memory capacity is also limited. However, they are least susceptible to performance degradation from metals and liquids.13.56 MHz is the typical frequency of use for HF RFID system. Compared to LF, HF tags are less expensive than LF tags. They also offer a fair performance in the presence of metals and liquids, HF RFIDS are currently the most widely available systems. The RFID system operating in these frequency ranges have the fastest data transfer rate between the tag and the reader. They are mostly meant for long range operation.

B. Network Coding Strategy

The network coding strategy used is as described in Fig. 3 and the notations used are described in Table 1.

The types of network coding are simple Ex-OR coding linear NC and random NC. This section applies a linear network coding procedure for distributed storage. The data captured in the lower level is from different nodes. The data is combined with some linear coefficients. There are 2-levels of linear coding as shown in Figure 4. The lower level combines the data from different tags and the level 2 coding combines the data from multiple readers. This network coding is at the

client node. The higher layer network coding occurs at the secondary node or at the server.

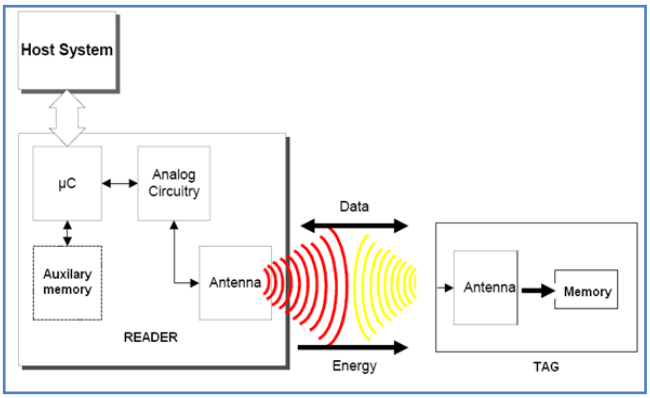


Figure 3. Overview if RFID System

Table 1 Notations of Network coding model

S. No	Notation	Description
1.	N	Number of tags
2.	M	Number of Readers
3.	K	Number of client nodes
4.	T_i	Tag ids; i = 1-N
5.	R_j	Reader ids; j = 1-M
6.	C_k	Client nodes; k = 1- K
7.	{a,b,c}	Encoding coefficients –Level 1 Primary nodes
8.	{p,q,r}	Encoding coefficients –Level 2 Primary nodes
9.	{α, β, γ}	Encoding coefficients Secondary nodes

The layered approach followed is described here. The data is captured through a .net application and the solution pushes the data to be combined into a network coding middleware. The middleware operation is performed in the mat lab environment. Using export the data is added as an input for the mat lab. By down casting the network coded solution from mat lab is uploaded into the server.

The network coding middleware's abstracted view is shown in Figure 5.

C. Client Server Model

The RFID readers are connected with the clients by Wi-Fi network model. The client supports different mobile RFID readers. From the client the environment uses the local area network to connect the data to server. The hospital main server is visualized to get connected with many clients in the hospital environment.

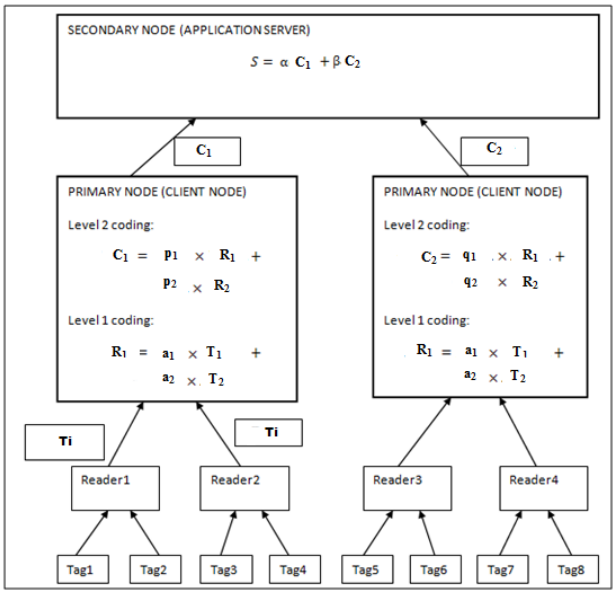


Figure 4. Architecture for Multilevel network coding

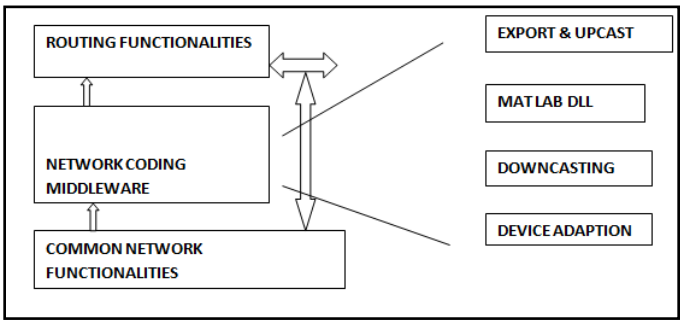


Figure 5. Network coding middleware

IV. Case Study

A. EHR Scenario

The idea of our proposal addresses a Pervasive Hospital environment from where the data is uploaded to the cloud server. Also the data can be downloaded from the cloud to the user community. The overall architecture is modelled as a 4-layer set up and is depicted in Figure 2. The topmost or fourth layer forms the data cloud for the social media. The data cloud upload community in our experiment comprise of Physicians, Nurses, Lab technicians etc. Whereas the download community includes the patients in the list and legal registration model is assumed for all the users. The data uploaded and downloaded is assumed to follow Health Level Seven (HL7) International format. The design of fourth layer of this overall architecture is beyond the focus of this work. Our work focuses on the experimental implementation of the first 3-layers. Layer 1 comprises of mobile RFID readers, layer 2 with client Personal Computers (PCs) and the 3rd layer with high end PCs acting as application servers. Our contributions in these layers are

1. RFID reader configuration and RFID healthcare application deployment in the first layer
2. Design and deployment of Network coding middleware in the primary and secondary nodes

(client PCs and Application PCs respectively) of the second and third layers

3. Integration of the three layers and testing of upload and download processes

B. Functional Blocks

Handheld, Passive RFID readers are used as data capturing devices. The readers can be used to capture healthcare informations like Nursing details of In-ward patients, Observation reports of a physician, Lab report inferences, Hospital asset tracking etc., The modules tested in our work comprises of observation details of in-ward patients by the nurse and physician. The readers of mobile types are used. The devices follow passive EPC gen 2 protocols for tag communication. The captured data is exported to the second layer. The user module based applications are developed and deployed into the readers. Suitable admin module leverages the authenticated users to access their module. The authenticated users can import the details of the patients privileged to their limit. They can modify and export relevant observations about the patient.

The experiment uses Motorola's MC9090-G RFID handheld reader. The MC9090-G is a rugged mobile computer from Motorola which provides mobile workers with a flexible, always-on data connection to critical applications and systems. Equipped with the latest advances in mobile technology, the MC9090-G provides support for the richest enterprise applications, empowering mobile workers to capture and access critical and emergency information in real time. The MC9090-G offers the latest Intel processor designed to handle the specific demands of mobility, as well as robust persistent storage capabilities and multiple advanced data capture options. A choice of the two most robust Microsoft operating systems —Windows Mobile 5.0 or Windows CE 5.0 — gives the flexibility to select either a familiar feature rich environment or a robust customizable application specific environment.

The Figure 6 expresses the overview of the simple RFID system for inward patient. Each patient is provided with the wrist band where RFID tag is attached to it. The tag contains unique serial number which is the tag ID of the patient. Depending upon the memory capacity the tag may contain information like name of the patient, disease, etc. The nurse or the physician carries the RFID readers in order to find the patient who needs to medicate or to be observed periodically under various conditions.

The reader is configured either with USB or Bluetooth protocol. The reader facilitates Wi-Fi facility also but our experiment does not use the provision. The readers are synchronized with the help of Active Sync. We designed our application in a kiosk mode in each reader. The reader applications of various modules like registration (Admin module), medication (Nurse Module), observation/monitoring (Physician module) are developed using Visual studio. The application programming uses .net framework 3.5 with a database of SQLCE. A model of the application User interfaces are designed in PC and deployed in to the reader.

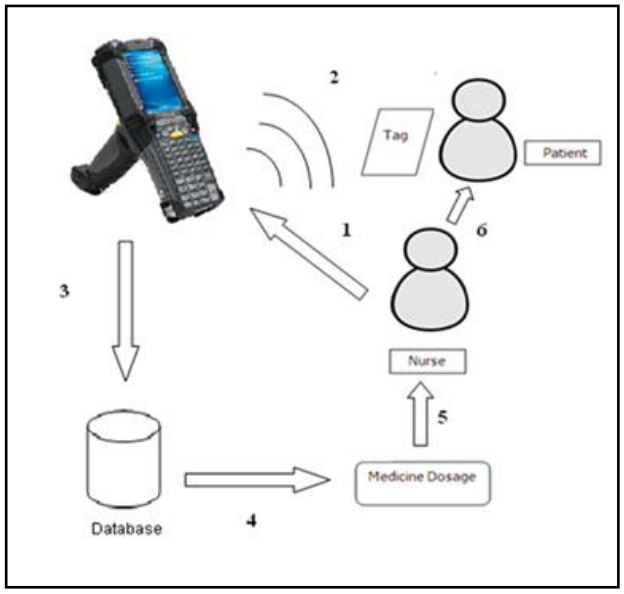


Figure 6. RFID application set up

The RFID readers capture the data from the tags and push them into the nodes of the upper layers. The contents comprise of different informations about the same patient like Medication and Lab inference or the informations about different patients. For the design of network coding middleware, we execute three sub-modules of design namely

Module 1: Modelling the network coding method

Module 2: Network coding architecture at the nodes

Module 3: Network Coding Middleware design

Module 1 of our design uses linear network coding. Each component of the information has its own encoding variable. The design follows the notations as listed in Table 3.1. Also Figure 4. depicts the model of linear coding followed in our work. Stage by stage content based encoding makes the retrieval of the contents comfortable. The primary nodes or the client nodes follow two level linear coding whereas the secondary nodes follow the single level linear coding. The complexity of the entire network coding processes depends upon the number of patients or tags, number of client systems and number of readers used.

Linear Coding at Client Nodes:

$$N = 2 \quad (1)$$

$$M = 2 \quad (2)$$

$$K = 2 \quad (3)$$

$$T_i = \{T_1 , T_2\} \quad (4)$$

$$R_j = \{R_1 , R_2\} \quad (5)$$

$$C_k = \{C_1 , C_2\} \quad (6)$$

$$R_1 = a_1 \text{ x } T_1 + a_2 \text{ x } T_2 \quad (7)$$

$$R_2 = b_1 \text{ x } T_1 + b_2 \text{ x } T_2 \quad (8)$$

$$C_1 = p_1 \times R_1 + b_2 \times T_2 \quad (9)$$

$$C_2 = q_1 \times R_1 + q_2 \times R_2 \quad (10)$$

Server side encoding

$$S = \alpha \times C_1 + \beta \times C_2 \quad (11)$$

Module 2 consolidates the modified architecture of the nodes. The nodes are introduced with a new virtual layer of network coding. This middleware layer of network coding is sandwiched between the network common functionalities and the routing layers. The process of uploading accompanies encoding and the process of downloading involves level wise decoding. The functionality flow in the architecture is as follows. The Ethernet connected client stores the collected information from the reader. After that, the node seeks for forwarding the information. The demo set up has a virtual network layer for this purpose. The virtual network layer functions are grouped into 3-sublayers as common, network coding and routing. The common functionalities identify the common category of the patients or reader or clients. Then the middleware of network coding performs the linear network coding. The coded results are routed to the server with the routing sub layer functions.

Module 3 of our design provides the insight of our network coding middleware. The demo version of our middleware makes use of mat lab for linear coding. The entire middleware is developed as an application using .net. The application collects the tag informations in a string format and downcast the format. These down casted inputs are exported to mat lab. The common random encoding matrix generation and matrix multiplication forms the basic logic of our mat lab code. We pre-processed the information in the application and size it for (7, 4) linear encode process. The mat lab code for linear encoding is compiled into a dynamic link library and patched with our .net application. The interoperability with mat lab is exploited in our demo version and the professional middleware can make use of server based format conversion and coding.

The informations collected by the reader are stored in the data base with suitable tables. After buffering them, the middleware of the virtual network layer search for the common tag id contents and the format conversion is executed. The information values are exported to Mat lab. The source code of linear net cod and the corresponding dlls contributed by Sadeghi, Shams and Traskovare (2010) are used as the basic blocks for our linear encoding. The scale down form of our code demo includes in the primary level an information matrix collected for a single patient. We assumed a 8*3 matrix, such that each column represents each module (Nurse module, Physician module and lab module), and the rows represent a 8-bit form of a data captured at the specific instance. This is a matrix assumed at a time stamp after one of the events and an observation period of 15-minutes. The encoding matrix is a binary matrix of size 3*1 at this level having each row value representing True or False about the event of each module occurrence. During decoding the request initiated by the corresponding module selection forms the encoding matrix. To avoid the code complexity in the demo simple data informations are coded and tested. The data storage and retrieval of network coded output is done with a local application server.

V. Output Samples

In the first stage of experiment, the application is deployed in the handheld reader. The .net based user interface as shown in Figure 7 allows the user to interact through the device. Synchronization of the device is essential and the output sample in Figure 8 reveals that. This invokes the application at the reader. The patient id is read as per sample in Figure 9. Figures 10, 11, 12 and 13 demonstrate the application sequence of medication.

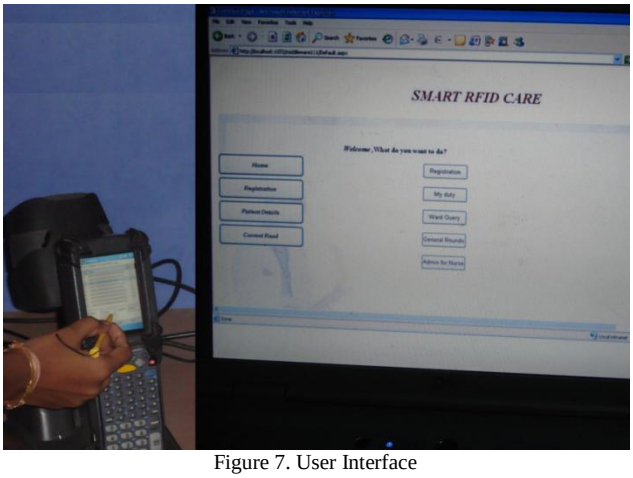


Figure 7. User Interface

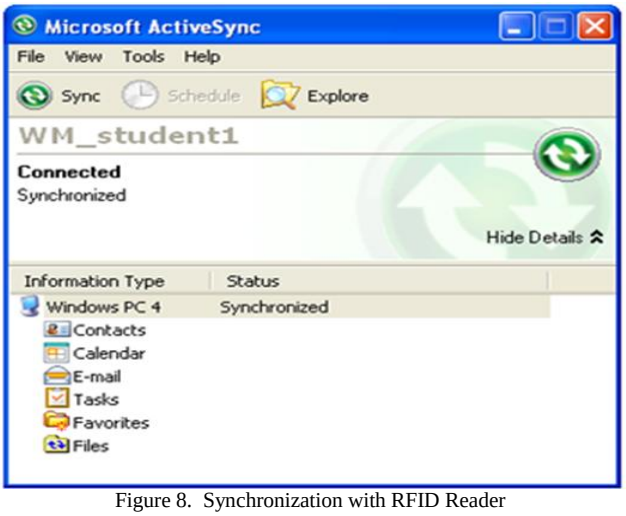


Figure 8. Synchronization with RFID Reader

Figure 9. Tag ID display

Figure 10. Registration Form

Figure 11. Patient details

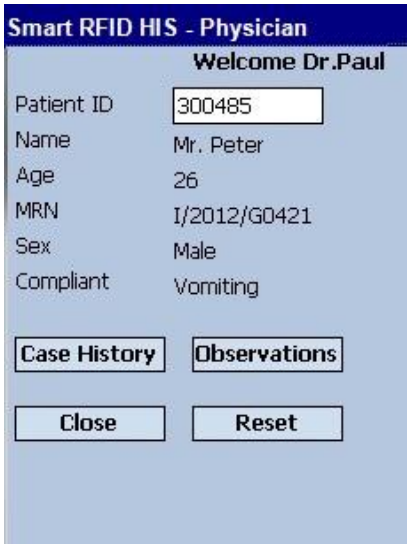


Figure 12. Case History

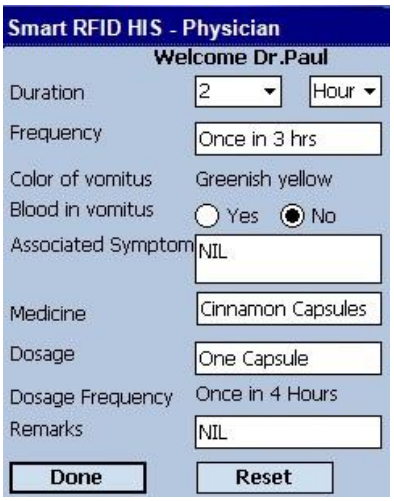


Figure 13. Medication details

VI. CONCLUSION

A simple linear network coding application is verified through our case study.The compatibility issues between the matlab computation of network coding and .net application code restrict our case study with limited sample of results.

The benefits of our design are (i) the server application is ignorant about the source of the contents collected at a specific time stamp (ii) while retrieving the data, the hand held device requests are very simple without specifying the details of their module; this saves the transmission cost (iii) In case of malfunction of any of the clients, the information can be retained with the history of encoding variables and the neighbour clients.

REFERENCES

[1] Christos Gkantsidis, Pablo Rodriguez Rodriguez, 'Network Coding for Large Scale Content distribution' , *IEEE Infocom 2005*

[2] Dimakis A.G, Prabhakaran.V, and Ramchandran.K, 'Ubiquitous Access to Distributed Data in Large-Scale Sensor Networks through Decentralized Erasure Codes' In *Proc. IEEE/ACM Int. Symposium on Information Processing in Sensor Networks (IPSN)*, April 2005

[3] Bo Chen, Reza Curtmola, Giuseppe Ateniese, Randal Burns, 'Remote Data Checking for Network Coding-based Distributed Storage Systems', *CCSW'10*, October 8, 2010, Chicago, Illinois, USA

[4] S. Rhea, C. Wells, P. Eaton, D. Geels, B. Zhao, H. Weatherspoon, and J. Kubiatowicz, "Maintenance-free global data storage," *IEEE Internet Computing*, pp. 40–49, September 2001.

[5] R. Bhagwan, K. Tati, Y.-C. Cheng, S. Savage, and G. M. Voelker, "Total recall: System support for automated availability management," in *NSDI*, 2004.

[6] Alexandros G. Dimakis, P. Brighten Godfrey, Yunnan Wu, Martin Wainwright and Kannan Ramchandran,' Network Coding for Distributed Storage Systems, IEEE Transactions on Information Theory, September 2010, **Volume:** 56 , Issue: 9 **Page(s):** 4539 - 4551

[7] [Aviv Shachak, Alejandro R. Jadad, 'Electronic Health Records in the Age of Social Networks and Global Telecommunications', *JAMA*. 2010; 303(5):452-453

[8] Sue Ansell, ' Health IT and EMR Predictions for 2015', *Canadian EMR*, January 2, 2011

[9] You Chen, 'Uncovering Anomalous Usage of Medical Records via Social Network Analysis', Presentation, 2, November, 2011; Presented at the *TRUST Autumn 2011 Conference*, Available at http://hiplab.mc.vanderbilt.edu/~ychen/trust.pdf

[10] Stanley Feld, 'What Are The Defective Assumptions Made For ACO Implementation?', *Repairing the Healthcare System [online]*, May 22, 2011, Available at http://stanleyfeldmdmace.typepad.com/repairing_the_healthcare_/2011/09/what-are-the-defective-assumptions-made-for-aco-implementation.html

[11] Taridzo Chomutare' , Luis Fernandez-Luque, Eirik Årsand, Gunnar Hartvigsen, 'Features of Mobile Diabetes Applications: Review of the Literature and Analysis of Current Applications Compared Against Evidence-Based Guidelines', Journal of Medical Internet Research, Vol 13, No 3 (2011).

A Secured Communication Based On Knowledge Engineering Technique

M. W. Youssef
Head of Computing & Information Division
The Shoura Assembly, Cairo, Egypt

Hazem El-Gendy
CS Dept. Chair, Faculty of CS & IT,
Ahram Canadian University

***Abstract*— Communication security has become the keynote of the "e" world. Industries like eComm, eGov were built on the technology of computer networks. Those industries cannot afford security breaches. This paper presents a methodology of securing computer communication based on identifying typical communication behavior of each system user based on the dominant set of protocols utilized between the network nodes.**

Keywords- Computer Communications; Computer/ communications Protocols; Network Security; Authentication; Scrambling; Encryption; Standard Protocols; ISO Open System Interconnections (OSI) Model; object behavior analysis; knowledge engineering.

I. INTRODUCTION

Modern Industries such as *eCom* (electronic commerce) and *eGov* (electronic Government) [1] depend heavily on computer networks security [1-16]. Security has been managed in many ways such as: authentication [2], securing networks communication using encryption and scrambling [3], protocol modification [4] and others used firewalls [5].

The TCP/IP has become the industry standard for computer networks protocols. The TCP/IP depends on the OSI model which consists of seven layers covering from physical layer to application layer. Each one of those layers utilizes several protocols in order to perform its function. When computer nodes communicate over the network, they use those protocols relevant to each layer.

Accordingly, various types of protocols services can be utilized in a network environment. HTTP pages, HTTPS, SNMP, FTP, SMTP, POP3, IMAP, DNS, SSH, TELNET, SSL, TCP, ICMP, SIP, UDP, Media Streaming and a range of other ports with a variety of services [6]. Each one of the previous protocols is associated with what is called a port number [7].

A port number is a way to identify a specific process to which a network message is to be forwarded when it arrives at a server. For the Transmission Control Protocol and the User Datagram Protocol, a port number is a 16-bit integer that is put in the header appended to a message unit. This port number is passed logically between server transport layers and physically between the transport layer and the Internet Protocol layer and forwarded on. Some services or processes have conventionally assigned permanent port numbers. These are known as well-known port numbers. In other cases, a port number is assigned temporarily (for the duration of the request and its completion) from a range of assigned port numbers. This is called an ephemeral port number.

The main hypothesis of this research is, each network user has a network utilization pattern, from that pattern, a user behavior can be detected then a user profile can be deduced. For example, some users tend to use only HTTP protocol even for their emails; others use SMTP protocol for emails, FTP for file transfer and HTTP for WEB sessions.

In order to do that a knowledge base [8] is created containing users behavior models. The knowledge base is generated from captured protocols generated from each session. That requires a Packet filtering mechanism [9] that is able to capture the traffic, identify traded protocols, apply statistical analysis process on those protocols [10], generate statistics that represent utilization patterns and store deuced profiles in the system knowledge base. Eventually, behavioral analysis programs run against the stored information in the knowledge base to generate a user network utilization behavior model.

The system is self-learning and has the ability to tune itself with every additional acquired bit of information. Accordingly, the system starts with zero knowledge, but as users establish more sessions its knowledge increase and overtime tune itself until it reaches a stable user behavior model.

After having a stable user behavior model, the system watches for behavior anomalies, as with each user session the system keeps an eye on users' behavior and comparing those behaviors to the existing knowledge base, for any reason, if a user behavior changed (anomaly detected), the system gives alerts to the system administrator that a user is doing something that is different from its normal behavior. The system administrator can then investigate the problem.

In this system, users are identified by their MAC addresses and protocols are identified from protocol identifier in packet headers at one stage and from port number at another.

The rest of this paper is organized as follows. Section two presents the structure of packet headers. Section three presents the relationship between protocols and port numbers. Section four discusses packet filtering techniques. Section five presents the design of the knowledge engineering security system and finally, section six is the paper conclusion and proposed future work.

II. TCP/IP HEADER LAYOUTS

The TCP/IP header consists of several fields, each of which has a specific function in data communication. The layout of the TCP/IP header is presented in figure 1. An important field in TCP/IP header is protocol identifier field [29]. This field will be used to build the system knowledge base.

The TCP/IP header Layout

00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
Version				IHL				Differentiated Services								Total length															
Identification																Flags			Fragment offset											
TTL								Protocol								Header checksum															
Source IP address																															
Destination IP address																															
Options and padding :::																															

Figure 1: The TCP/IP header Layout

In the TCP/IP, every packet header contains some important information. The main information is:

- IP source address.
- IP destination address.
- Protocol (whether the packet is a TCP, UDP, or ICMP packet).
- TCP or UDP source port.
- TCP or UDP destination port.
- ICMP message type.

In addition, packet headers contain additional information regarding packets that aren't reflected in the packet headers, such as:

- The interface the packet arrives on.
- The interface the packet will go out on.

In this research, packet headers will be intercepted and protocol identifiers will be captured to be used for building the behavioural model in the knowledge base.

III. PROTOCOLS AND PORT NUMBERS

For each application-specific or process-specific protocol there is a port number serving as a communications endpoint. It is used by the transport protocols of the Internet Protocol Suite, such as Transmission Control Protocol (TCP) and User Datagram Protocol (UDP). Servers for each protocol service use that given port number, that port number allows the identification of running protocol and types of connections simply by specifying the appropriate port number. Port numbers are divided into three ranges: well-known ports, the registered ports, and the dynamic or private ports. The well-known ports are those from 0 through 1023. Examples include: :File Transfer Protocol (FTP) port 21; Secure Shell (SSH) port 22; Telnet remote login service port 23; Simple Mail Transfer Protocol (SMTP) port 25; Domain Name System (DNS) service port 53; Hypertext Transfer Protocol (HTTP) used in the World Wide Web port 80; and so on. The registered ports are those from 1024 through 49151. IANA maintains the official list.[1] The dynamic or private ports are those from 49152 through 65535. One common use is for ephemeral ports.

Some examples of the ways in which port numbers are to selectively route packets to or from your site:

- Identify all incoming SMTP connections.
- Identify email and FTP services
- Identify dangerous services like TFTP, the X Window System, RPC, and the "r" services (*rlogin*, *rsh*, *rcp*, etc.).
- Identify all connections to or from certain systems.

In this research, port numbers will be used to identify services and applications protocols to be used for building the behavioural model in the knowledge base.

IV. PACKET FILTERING TECHNIQUES

Packet filtering systems route packets between internal and external hosts, in some cases, they do it selectively. They allow or block certain types of packets in a way that reflects a site's own policy. There are several types of packet filters, the most common are:

1. A screening router.
2. Proxy Services.
3. Using a Combination of Techniques and Technologies.

A. *A Screening Router To Do Packet Filtering.*

This type of routers is used in a packet filtering firewall which is known as a *screening router*. A screening router looks at packets more closely and in addition to determining whether or not it *can* route a packet towards its destination, a screening router also determines whether or not it *should*. "Should" or "should not" are determined by the site's security policy, which the screening router has been configured to enforce. [2]

B. *Proxy Services.*

Proxy services are specialized application or server programs that run on a firewall host: either a dual-homed host with an interface on the internal network and one on the external network, or some other bastion host that has access to the Internet and is accessible from the internal machines. These programs take users' requests for Internet services (such as FTP and Telnet) and forward them, as appropriate according to the site's security policy, to the actual services. The proxies provide replacement connections and act as gateways to the services. For this reason, proxies are sometimes known as *application-level gateways*.[3]

C. Using A Combination Of Techniques And Technologies.

Some protocols such as Telnet and SMTP can be more effectively handled with packet filtering. Others such as FTP, Archie, Gopher, and WWW are more effectively handled with proxies.

On the other hand, Proxy services are effective only when they're used in conjunction with a mechanism that restricts direct communications between the internal and external hosts. Dual-homed hosts and packet filtering are two such mechanisms. If internal hosts are able to communicate directly with external hosts, there is no need to use proxy services.

Accordingly, this research has used a combination of techniques in order to collect the required packets.

V. The Design Of The Knowledge Engineering Security System

Knowledge engineering as a tool has been used in this research due to its nature of being data driven rather than instruction driven which is required in this research state where data is the main actor in the system. A similar technique was used in [11, 12]. The presented system consists of two parts those parts are:

1) *The Knowledge Engine.*
2) *Anomalies Detector.*

The two modules are working continuously in conjunction in. The main design of the knowledge engine is presented in figure 2.

A. Design of the Knowledge Engine

The Knowledge engine components are:

1. Packet capturing module
2. Collected Packets Base.
3. Packets Classifier.
4. Classified Packets Base.
5. Behaviour analyser module (Knowledge Engine).
6. Knowledge Base (Users Behaviour Models).

1) Packet Capturing Module.

This main function of this module is to capture packets coming and going through the network. This module is used for both the system knowledge engine and the anomalies detector. For the first it stores the collected packets in the knowledge engine collected packets base and for the second it stores the collected packets in the anomalies detector collected packets base.

2) Collected Packets Base.

It contains the knowledge engine collected packets.

3) Packets Classifier.

This module is responsible for analysing packets at network layer to extract certain key information according to a set of criteria that a packet must match to be accepted in a trace buffer. There are two things used to make that sub module work:

Packet offset: it is used to define a location based on the start of the packet. [13]

Protocol offset: it takes into account that the packet analyser may be applied on networks that use different frame types. Its offset values are based on the start of data after the MAC header. [13]

An example of data offsets is presented in table 1.

By the end of the process, it stores classified packets in the classified packets base.

4) Classified Packets Base.

It contain all the system classified packets, it represents the first step in the knowledge building process.

5) Behaviour analyser module (Knowledge Engine).

The main function of this module is modelling users' behaviours. That is done by applying two methods. The first method is used to model the typical behaviour using finite state machines. The second method is used to model the signature of behaviours.

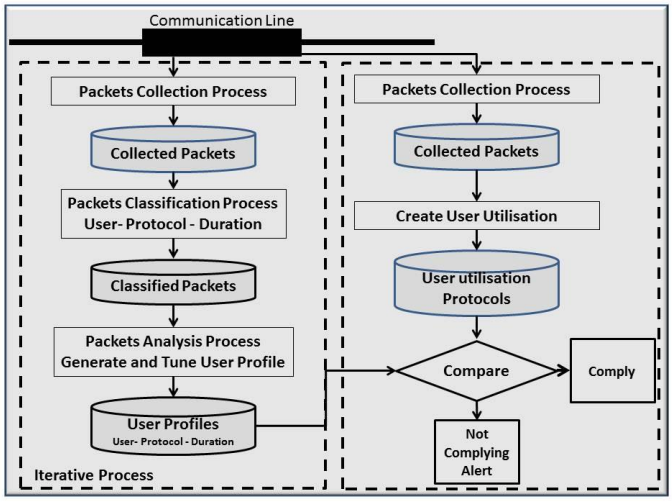


Figure 2: Design of the Knowledge Engineering Engine

Offset	Type	Value
0x00/0d	Packet (Ethernet)	Destination Hardware Address
0x06/6d	Packet (Ethernet)	Source Hardware Address
0x0C/12d	Protocol	Source IP Address
0x10/16d	Protocol	Destination IP Address
0x14/20d	Protocol	Source Port Number
0x16/22d	Protocol	Destination Port Number
0x21/33d	Protocol	Flags fields
0x1C/28d	Protocol	Data over UDP
0x28/40d	Protocol	Data over TCP

Table 1 Summarised data offset

Each module generates its specific instances to form user behaviour. Its attributes are:

- BehaviourNo: it contains an identifier value of a specific behaviour.
- BehaviourName: it contains name of the intended behaviour.
- BehaviourType: it contains type of the intended behaviour
- BehaviourPhase: it contains phase of the intended behaviour
- InitialCertainty: it contains the degree of certainty of the first instance of the intended behaviour.
- CertaintyIncrement: it contains the increment value added to the degree of certainty at each occurrence of the intended behaviour.

6) Knowledge Base (Users Behaviour Models).

All behavioural models are presented in the system knowledge base.

B. Design Anomalies Detector.

The anomalies detector components are:

1. Packet capturing module.
2. Collected Packets Base.
3. Session utilisation module:
    - Packets Classifier.
    - Classified Packets Base.
    - User utilisation profile generator (per session analyser module).
4. Anomalies detector.

1) Packet Capturing Module, Collected Packets Base and Session Utilisation Module.

The main objective of this phase is to build on the run spontaneous user behaviour to compare it the existing behaviours in the system knowledge base. Accordingly, in this phase, stages 1, 2 and 3 are very similar to what takes place in phase one, the knowledge engine phase, the difference is, it is done session by session or flow by flow based in the type of the captured protocol.

2) Anomalies Detector.

This module is responsible for detecting user behaviours anomalies and deducing actions towards detected behaviours. The anomalies detector module consists of two sub modules:

the rule base for behaviour comparison and the inference engine to decide actions against that behaviour. Both sub modules use a fact base of states of behaviour.

a) Inference engine rules base

This rule base is used to store already acquired knowledge in the domain of anomalies rather than the ordinary domain. This knowledge is represented through action (situation) and premise (conclusion) production rules.

b) Inference engine

This sub module makes forward-chaining (event driven) reasoning on events in the fact base and heuristics in the rule base to deduce consequences of detected behaviours and then their appropriate reactions to broadcast that to the system administrator. In addition, in ambiguous behaviour, the inference engine supports both of uncertainty and conflict resolution.

c) Fact base of States of Behaviours.

This fact base is used to store information on behaviours that are in progress or completely detected. Its attributes are:

- InstanceNo: it contains an identifier value to a specific behaviour instance.
- PacketNo: it contians an identifier value accompanied with each inspected packet.
- BehaviourNo: it contains an identifier value to the intended behaviour.
- CurrentCertainty: it is assigned with the degree of certainty of the intended behaviour instance.

VI. Conclusions And Future Work

This paper presented discussion of the design of a proposed network security technique that is based on knowledge engineering concepts. The system utilises a packet filtering system to build a behavioural model of users' behaviour and detects anomalies in those behaviours.

In order to do that the system utilised an inference engine that is based on the frequent reevaluation of the stored states of users' behaviors. The advantage of that approach stems from relaying on a dynamic control structure rather than a static control structure. Accordingly, the system works in data-driven way. The system rules can communicate with one another only by way of the data.

The system utilized a “stateful” packet filtering engine. That technique allowed the system to perform the four basic networks security processes: alert, protect, respond and manage.

An integral part of the system is its knowledge base. That knowledge base allowed the system to detect anomalies in users' behaviour and generate alerts. More importantly, the system has the ability to learn over time to tune and modify its models.

Finally, the system needs to be tested with large volume of data in order to measure its performance and time considerations. Also, the system needs further tests where the computation of the conflict set is a non-trivial problem.

REFERENCES

[1] H. Makhlouf, M. W. Youssef and S. Ismaeel, "التجارة الالكترونية و الحكومة الالكترونية", Ain Shams Press Inc, second edition, 2008

[2] M. W. Youssef and H. Elgendy, "Applying Open Networks Communication Authentication By Scrambling and Encrypting Layer Two Packet Content.", IJCSNS, June 2011.

[3] M. W. Youssef and H. Elgendy, "Applying Open Networks Communication Authentication"; IEEE conference Security and Applications, St. Petersburg, 2011.

[4] M. W. Youssef, "Securing Computer Networks Communication By Modifying Computer Network Communication Protocols"; IEEE conference Security and Applications, St. Petersburg, 2011.

[5] O. I. Sherif, "A Firewall Based Schema for Computer Network Security", M.Sc. thesis, Institute of statistical studies and research, 2001.

[6] Comer 2000, Sect. 11.2 - The Need For Multiple Protocols, p. 177, introduces the decomposition in layers and Sect. 11.3 - The Conceptual Layers Of Protocol Software, p. 178.

[7] "Port Numbers". The Internet Assigned Numbers Authority (IANA).

[8] Avelino J. Gonzaie, Douglas D. Donkel, "The Engineering of Knowledge-based Systems Theory and Practice", Prentice-Hall, Inc. 1993.

[9] Youssef M. W. and Roshdy K., "A Proposed Packet Filtering System To Protect Open Networks", ISCA, 2005.

[10] Ronald E. Walpole, Raymond H. Myers, "Probability and statistics for engineers and scientists", Macmillan publishing company, 2001.

[11] Youssef M. W. and G. Morris, "A Decision Support System for Forecasting Dynamic Objects Behaviour And it's Application on Lake Qaroun", Port Said Engineering Research Journal, March 2003.

[12] Youssef M. W. and G. Morris, "Using Multiple Regression Models To Simulate Dynamic Objects Behaviour In A Decision Making System", Port Said Engineering Research Journal, September 2003.

[13] J. Casad, "Teach yourself TCP/IP in 24 Hours", SAMS, Second Edition, 2001.

[14] [14] Nabil El Kadhi and Hazem El-Gendy, "An Intelligent Bidirectional Authentication Method", International Journal of Computer and Network Security, Vol. 2, No. 10, 2010, pp. 1-7.

[15] Mohamed Wagdi and Hazem El-Gendy, "Applying Open Networks Communications Authentication", Proceedings of the 11th International Conference on Intelligent Transport System – Telecommunications sponsored by IEEE and IEEE Communications Society, Saint. Petersburg, Russia, 23-25 Aug. 2011, pp. 650-657.

[16] Mohamed Wagdi Youssef and Hazem El-Gendy, "Scrambling and Encrypting-Based Authentication for Open Networks Communications", International Journal of Computer Science and Network Security, June 2011, pp. 24-29.

Enhanced Authentication Mechanisms for Desktop Platform and Smart Phones

Dina EL Menshawy, Hoda M. O. Mokhtar, Osman Hegazy
Information Systems Department
Faculty of Computers and Information, Cairo University
Cairo, Egypt

***Abstract*— With hundreds of millions using computers and mobile devices all over the globe, these devices have an established position in modern society. Nevertheless, most of these devices use weak authentication techniques with passwords and PINs which can be easily hacked. Thus, stronger identification is needed to ensure data security and privacy. In this paper, we will explain the employment of biometrics to computer and mobile platforms. In addition, the possibility of using keystroke and mouse dynamics for computer authentication is being checked. Finally, we propose an authentication scheme for smart phones that shows positive results.**

Keywords- Biometrics; Keystroke; Mouse; Authentication; Smart Phones; Touch Screens; Touch Pressure; Touch Contact Size.

I. INTRODUCTION

Today we are witnessing a tremendous increase in the use of computers and smart phones for storing sensitive information and accessing on-line services. These devices have become important tools in many people's daily activities, and are consequently used for many purposes including: communication, entertainment, storing confidential personal and business information. Therefore, the hacking of a computer or a mobile device can have negative implications like the invasion of privacy, the opportunity to impersonate user, and even severe financial loss. Current user authentication for computers and mobile phones is provided by the personal identification numbers (PIN) and passwords which have a number of inherent weaknesses such as the ease of figuring out one's PIN. In general, there are three levels of computer security mechanisms: the first mechanism depends on something a person carries, such as an ID badge with a photograph, while the second scheme relies on something a person knows, such as a password. Finally, the third approach is related to a person's human attributes, such as fingerprint and/or signature [1]. The increasing need for improving security systems led to more research in the application of biometrics in authentication systems. The term biometrics originates from the Greek words bios (life) and metrikos (measure). Biometrics refers to the identification of a person based on his/her physiological or behavioral characteristics. People have personal characteristics that uniquely identify them such as hand signature, fingerprint and voice. In general, biometrics is mainly divided into two categories, namely, physiological biometrics and behavioral biometrics. *Physiological biometrics* identifies a person based on his/her physiological characteristics such as eye retina, whereas *behavioral biometrics* relies on detecting the behavioral attributes of the user, such as keystroke dynamics [2]. Biometrics became popularly used as a tool for security because of its universality and distinctiveness. Mainly, there are two capabilities of biometrics which are identification and verification. Identification is the process of determining a person's identity; whereas verification ensures that the person requesting the access is the one he claims to be [3]. A biometric system consists of several modules, the main components are: a sensor module, a feature extraction module, and a classification module. The sensor module captures the trait, and then the feature extraction module extracts a feature set from the captured data. After that, the classification module compares the extracted feature set with reference feature sets to validate a claimed identity [1]. Finally, a biometric based authentication system can be evaluated using either a genuine test or an impostor test, described as follows:

- The genuine test (or False Rejection Rate (FRR)): when the user enters an input that is far away from his own template.
- The impostor test (or False Acceptance Rate (FAR)): when the user enters an input that is very similar to another user's template [1].

The main contribution of this work can be summarized as follows:

1. Explored the use of machine learning techniques in biometric authentication for both desktop application and smart phones.
2. For desktop platform, we examined the employment of neural networks and k-means clustering with focus on only two types of behavioral biometrics; keystroke and mouse data.
3. Also, we constructed a multi-modal biometric system based on both keyboard and mouse data. Fusion at the feature level was examined for a better degree of accuracy. Feature level was accomplished by merging both forms of data to create a new behavioral feature vector.
4. For smart phones, we proposed an authentication mechanism based on different metrics. Again, neural networks and k-means clustering were implemented on the collected data. The investigations examined

other behavioral biometrics which is: key hold times, latencies, finger pressure and finger contact size.

The rest of the paper is organized as follows: we present in Section 2 a comparison between uni-modal and multimodal biometric systems. Section 3 discusses related work to the study of biometrics for computer and mobile phones authentication. Section 4 introduces the behavioral biometrics and discusses the types of data that will be used in the experiments. Section 5 presents the machine learning techniques used in our research. Section 6 describes our experimental approach and results. Finally, Section 8 concludes the paper and Section 9 presents directions for future work.

II. Multi-Modal Biometric Systems

Biometric systems depending on a single source of information are called uni-modal systems, while systems depending on multiple resources are named multi-modal systems. Sometimes, uni-modal biometric systems do not attain the required performance because they are more susceptible to problems such as noisy data, intra-class variations, inter-class similarities, non-universality and spoofing. On the other hand, multi-modal biometric systems handle some of these problems. They solve the problem of non-universality as multiple traits will guarantee adequate population coverage. Moreover, they overcome the spoofing problem as it won't be easy for an imposter to spoof multiple biometric traits of a real user [4]. Fusion can be achieved by applying multi-modal systems, fusion refers to combining or making use of multiple biometric traits to enhance the classification accuracy. There are three various levels of fusion: the first type is at the feature level where feature sets arising from several sensors are fused. The second kind of fusion is at the match score level where the scores produced by classifiers related to different biometric traits are aggregated. The third level of fusion is achieved at the decision level where the final outputs of multiple classifiers are merged through certain techniques. In our research, we will build a multi-modal biometric system then we will apply the feature level fusion on behavioral biometrics.

In this paper, we investigate a crucial problem in biometric data, namely, mining biometric data. Data mining has become an increasingly popular activity in all areas of research, from business to science, and currently, in biometrics. Biometric verification is gaining more attraction because most of the systems based on it are easy to incorporate in ordinary computer use and without user interaction. Besides, they do not need extra devices for authentication. The challenge is to integrate machine learning techniques into biometrics verification leading to the evolution of the term biometric data mining. Hence, biometric data mining (BDM) is the application of knowledge discovery techniques to biometric information with the purpose of identifying underlying patterns [5].

III. Related Work

Recently a number of researches were conducted to explore the utilization of machine learning techniques in different biometric systems. Several works on keystroke biometrics have already adopted approaches based on different metrics, sampling methodologies and data analysis techniques. In [6], the authors proposed a benchmark testing suite composed of a database and a software that are publicly available for the research community to evaluate keystroke dynamics based systems. The software offers several functionalities, for example, it records timing information when a user enters a certain password. Also, the tool offers different types of keyboards to test typing evolution depending on this parameter. In [7], the authors designed, tested, and evaluated four different metrics related to keystroke analysis. The four metrics were key press duration, key press and release comparisons, relative keystroke speeds, and a metric based on shift key usage patterns. Each user typed the sentence"A quick brown fox jumps over the lazy dog" up to eleven times and the different metrics were recorded. In [8], the authors presented a design and an implementation of a remote authentication framework called TUBA for monitoring a user's keystroke-dynamics patterns and identifying intruders. They evaluated the robustness of TUBA through comprehensive experimental evaluation including two series of simulated bots. It was concluded that TUBA can be integrated with other anomaly detection systems to achieve remote monitoring and diagnosis of hosts with high assurance. In [9], the authors collected keystroke data in the form of digraphs when users enter a specific password, then rough sets were used to detect patterns in the typing rhythm. The analysis produced a sensitivity of 96%, specificity of 93% and an overall accuracy of 95%. On the other hand, some studies were done on using mouse movements as a biometric for authentication.

In [10], a user was asked to join the dots appearing on the screen, and then in the verification phase, the user should move the mouse in the same pattern as done in the enrollment step to check his/her identity. The testing was done in a classroom with students in the age group of 22-30. The error rate for this system was 20%. In addition, another study on using mouse biometrics was conducted in [11]. The k-nearest neighbor method was used to identify unknown mouse profile from a set of known user profiles; and the Euclidean distance was used to discover the nearest neighbor. A success rate of 92% for the first choice of the nearest neighbor was reached. Matching the second choice was 88% and matching second and third choices together was 80%. In [12], the paper investigated the effectiveness of user authentication using keystroke dynamics-based authentication (KDA) on mobile devices. A keystroke dynamics-based authentication mechanism was proposed with artificial rhythms and tempo cues for mobile user authentication. The novelty detector classifier was built. Then, subjects were asked to perform enrollment, login, and even intrusion to other subjects' accounts.

In [13], the authors investigated the authentication of users based upon three interaction scenarios: entry of 11-digit telephone numbers, entry of 4-digit PIN, and entry of text messages. The discussion focused upon the concept of keystroke analysis for users' authentication. The findings revealed the technique to be promising for certain users with average error rates below 5%. In [14], an application for the Android mobile platform was developed to collect data on the way individuals draw lock patterns on a touchscreen. Using a Random Forest machine learning classifier this method achieved an average Equal Error Rate (EER) of approximately 10.39%. In [15], six distinguishing keystroke features were used for user identification in smart phones. They optimized the front-end fuzzy classifier using Particle Swarm Optimizer (PSO) and Genetic Algorithms (GA) as back-end dynamic optimizers to adapt to variations in usage patterns. Finally, they provided a novel keystroke dynamics based PIN verification mode to ensure information security on smart phones.

IV. Behavioral Biometrics

Behavioral biometrics refers to a subset of biometrics which has to do with a person's behavior. Examples include keystroke dynamics, signature verification, and voice. Behavioral biometrics works on the characteristics that are developed naturally over time [2]. For instance, in keystroke dynamics, some features can be measured, for example, the typing speed, and the time taken between consecutive keystrokes. In the following discussion we will further elaborate two biometric types that will be later used in our experiments. Keystroke and mouse data will be shown in Sections 5.1 and 5.2 respectively. Also, the metrics that will be used in smart phones authentication will be presented. Finger pressure and finger contact area were considered distinguishable features across users which will be explored in Section 5.3.

A. Keystroke Dynamics

Keystroke dynamics means the pattern in which a user types characters, or numbers on a keyboard. Keystroke dynamics is used to define the person's identity because it resembles an individual's handwriting or signature. A user's keystroke rhythms are measured to generate a distinctive prototype of the user's typing patterns for use in authentication. One key advantage of using keystroke dynamics is that FRR and FAR can be fine-tuned by altering the acceptance threshold at the individual level. Also, keystroke movements can be captured constantly. Moreover, no additional hardware is needed to collect keystroke data, the keyboard is enough. Each user has a unique time for depressing and holding keys, as some people type certain words or characters faster than others. A lot of features can be derived from keyboard typing such as: duration time (the time of a key press), and latency (the time between"key up" and the next "key down") [16]. In this paper, different keyboard features are used to explore various attributes that are not common, the features are:

- The difference between two press events (PP).
- The difference between two release events (RR).
- The difference between one press and one release events (PR).
- The difference between one release and one press events (RP).
- The time to type the password: the total time taken to write a certain word or password [6].

B. Mouse Dynamics

Mouse dynamics is a recent behavioral biometric that is being used in authentication systems. Mouse dynamics means monitoring the users' activities through a human computer interface. It has been proven that user-based mouse movements can model the user's behavior. Features used to explain the users' behavior include drag and drop, click, and any other mouse movement. Moreover, we can compute some calculations such as the speed of moving the mouse across a certain distance. The key plus of using mouse dynamics to validate the identity of the user is that it does not need additional hardware to capture the users' behavior [17]. In addition, mouse dynamics is useful for continuous authentication since the user identity can be confirmed through the repeated mouse movements. In this paper, we focus on the following features to monitor mouse movements:

- Total time [T_{Total}]: time from when mouse button was first depressed to draw first segment until last segment was completed.
- Actual drawing time [T_{Draw}]: time excluding pauses when mouse button is released.
- Length [$Length_{tot}$]: total length of all drawn segments [18].

C. Finger Pressure and Finger Contact Size

We noticed that users of touch screens enter data in a characteristic manner as they exert different pressures and varying finger touch area. Therefore, we utilize both finger pressure and contact area as new biometric features for smart phones authentication. Analogous to keystroke studies but for smart phones with touch screens, two distinguishing features will be captured for all users:

- Finger pressure: the pressure of finger touch on the screen.
- Touch size: the contact area or the area pressed of the finger on the touch screen [19].

V. Biometric Based Authentication Techniques

In this paper, we employed neural networks and k-means clustering to study the accuracy and efficiency of behavioral biometrics for authentication in both desktop application and smart phones.

A. Neural Networks

A multilayer feed-forward neural network consists of an input layer, one hidden layer, and an output layer. In the hidden layer, each neuron performs a weighted summation of the inputs, which then passes a nonlinear activation function. The network output is formed by another weighted summation of the outputs of the neurons in the hidden layer. This summation on the output is called the output layer [20].

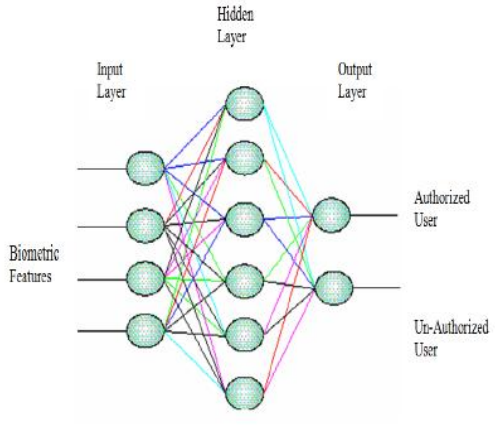


Figure 1: Multilayer Feed-forward Neural Network

B. K-means Clustering

The second machine learning technique used in this work is the traditional k-means clustering. Following the kmeans, the number of clusters is equal to the number of users. The main goal is to ensure that each user is correctly classified into a *single* cluster. When a user is classified into more than one cluster which will result in an in-accurate authentication, experiments can be repeated to ensure that the behavioral pattern does not change. K observations from the samples were selected at random as initial cluster centroid positions. Centroids were updated until they reached stable centers of clusters. A feature vector is created for each user. The aim is to partition N feature vectors into K disjoint subsets containing *Nj* feature vectors so as to minimize the sum-of-squares criterion [20]. Keystroke and mouse data denote the feature vector of a user in desktop platform while the finger pressure and finger contact size denote the feature vector of a user in smart phones.

VI. Experimental Setup And Results

The initial motivation for our research arose from the need to provide secure and unobtrusive methods for authenticating users of computers and mobile devices. The main objective is to have both low FRR and FAR as well as to achieve both high usability and high security of the system. In our research, the use of neural networks and k-means clustering will be investigated to mine behavioral biometric data and discover hidden features that help to increase verification accuracy.

Actually, mouse dynamics is used in GUI base applications, whereas a keyboard is essential for command line base applications so they are two related tools when dealing with computers. As a result, we will explore fusing those kinds of biometric data to construct an accurate multi-modal system. Also, a proposed scheme for smart phones authentication will be presented, the authentication will be based on finger touch pressure and finger contact size. The experimental details will be shown below:

A. Study of Behavioral Biometrics for Desktop Platform

In our proposed methodology, there are four important stages involved in keystroke and mouse dynamics based authentication system.

1. First, a user enrolls his/her feature vector.
2. Second, a preprocessing phase is done.
3. Third, neural networks were implemented using the feature vectors for each biometric trait on its own, then were applied on both keyboard and mouse data together. Here, we built three neural networks: the first one was based on keystroke data only, the second was based on mouse dynamics only. Last but not least, the third network worked on both forms of data together. Moreover, K-means clustering was implemented on the fused data.
4. Forth, the performance of the proposed system is evaluated.

The following sections clearly illustrate the experimental details:

1) Enrollment

We have conducted some experiments involving 20 participants, and collected experimental data over 3 weeks. Ages range from 18 to 30, and both males and females participants were involved to cover different ages and both genders along with different computer literacy or experience. A strong password was chosen containing capital and small letters, and numbers were used in the enrollment stage. All users were allowed to enter the same password several times to get used to typing it. Then, in the enrollment stage, each user typed the word ”DI19na25” twenty times and the keystroke features were recorded. Concerning mouse dynamics data, the 20 users were allowed to draw a line between 2 points and the mouse features were recorded.

Most literature work applied different machine learning technique on two keystroke features: duration of key hold and latencies. Here, we utilized a little bit features that may produce better results than those two features. We used PP, RR, PR, RP and total time to write a password. The definitions of those features were shown in a previous section. For example, when a user enters”DI19na25”, the four latency timings (PP, RR, PR, RP) were recorded for each pair of characters. Also, the total duration of writing the whole password is recorded. Concerning the mouse features, the following table clearly illustrates the used features:

Table 1: Mouse Features

ID	T_{Total}	T_{draw}	$Length_{tot}$
1	3.09	3.09	2008.6
1	3.09	3.09	2008.6
1	13.27	5.09	4016.5
1	27.92	7.68	6044.8

2) Preprocessing

As typing pattern of the same user varies from time to time although it is relatively unique for each user, normalization was done to improve the accuracy of classification. The normalization technique used was min-max as it has been shown to give good results. The preprocessing phase maps the feature vectors to fall into a small specified range. In the preprocessing phase, outliers were removed in order to improve the performance of the system [20].

3) Classification Using Neural Networks and Kmeans Clustering

In order to evaluate the feasibility of applying behavioral biometrics for authentication in secure systems, neural networks and K-means clustering were used in different experiments. We computed a profile for each member who will be later used as a reference in testing and evaluation. For both keyboard and mouse data, fifteen samples will be used for training and another five for testing. Neural networks were applied on keystroke and mouse data, each separately, and then fusion was done on the feature level. The feature vectors were classified using feed forward network. Feature level fusion is fulfilled by a concatenation of the feature sets acquired from several sensors. The main idea behind fusing more than one biometric trait is to improve the prediction rate. In this research, keyboard and mouse biometrics data are fused to form a single template. For example, if keystroke data is denoted by $\{X_1, X_2,...X_m\}$ while mouse data is expressed as $\{Y_1, Y_2,...Y_n\}$. The aim is to integrate both kinds of data to produce a new feature vector $Z = \{X_1, X_2,...X_m , Y_1, Y_2,...Y_n\}$ which better represents the user [21]. Three networks were built, the first one handled the keystroke data and the second network processed the mouse data. The fused feature vectors were fed into the third network. The networks were built in Matlab because it offers a great neural networks toolbox; also, it is relatively fast in testing.

4) Experimental Results

The three networks were run several times to compare the performance, it has been shown that fusing biometrics data achieves the best results, mouse data following it and keystroke attains the lowest accuracy. The numerical outputs are clearly presented in the following table:

Table 2: Recognition Accuracies of Three Networks

Data	Number of Characteristics	Accuracy (%)
Keystroke	33	54
Mouse	3	65
Fusion	36	72

FAR is computed as the percentage of imposters wrongly classified as legitimate users, and FRR is the percentage of legitimate users classified as imposters. In the testing phase, we divided the participants into 2 groups: a group of 10 representing authorized users, and the remaining 10 representing unauthorized users.

Also, FRR and FAR measures were calculated for each user then an average value was measured. The average FRR and FAR were 14 % and 17 % respectively. As fusion of both keystroke and mouse data gave promising results, k-means clustering was implemented to compare the performance with neural networks. The number of clusters was 20 as there were twenty participants, each cluster denotes a user. The whole samples were fed into the clustering algorithm. Twenty tests were run with random seeds, the tests resulted in an average accuracy of 79%.

B. Keystroke Dynamics for Smart Phones

Using Android 2.3.3 (API10) and Eclipse, we have developed a mobile application to collect data from different individuals about the way they type numbers on a smart phone. The handheld mobile device used in the experiment was a Samsung Galaxy Ace GT-S5830i with 832 MHz CPU and 158 MB memory.

The experimental approach is described as follows:

1. Let users practice typing PIN code until they can type smoothly.
2. Allow each user to type the PIN twenty times to create a database.
3. Classify the feature vectors using neural networks and clustering.
4. Test the biometric system using the error rates.

1) Enrollment

Twenty participants were enrolled in the experiment; each user was allowed to enter a PIN code. The PIN was chosen to be"9721" to avoid same horizontal and vertical alignment. The investigation required the participants to enter the PIN twenty times which will be used to create a reference profile. The first phase investigated the feasibility of authenticating mobile phone users based upon the traditional keystroke features used in computer authentication which are hold time of keys and latency between keystrokes. For example, when all users entered the PIN "9721", the time duration between the number pairs (9-7), (7-2) and (2-1) will be computed. Also, the duration time of holding each key will be recorded. In our case, each typed PIN consists of three latencies and four hold times features, resulting in seven features for each user. Again, min-max normalization is done in the pre-processing phase.

Since data entry on a standard numerical keypad on a PC differs from entering numerals on a mobile phone in terms of feel and layout, so a second phase of the study was implemented. This study seeked to evaluate the feasibility of using finger touch pressure and finger contact touch area as unique characteristics rhythms. Each user was allowed to practice by typing the PIN several times. In the enrollment phase, each user entered the PIN code 20 times to have a true profile for the typing pattern, and then finger pressures and contact area were recorded for each key. All erroneous trials were disregarded. Touch coordinates on button presses were supposed to be used in the experiments but when tried with different users, it has been shown that it cannot be used as a distinguishable feature. Screen location on touch screen of different users were similar so was not used in our investigations.

2) Classification Using Neural Networks and Kmeans Clustering

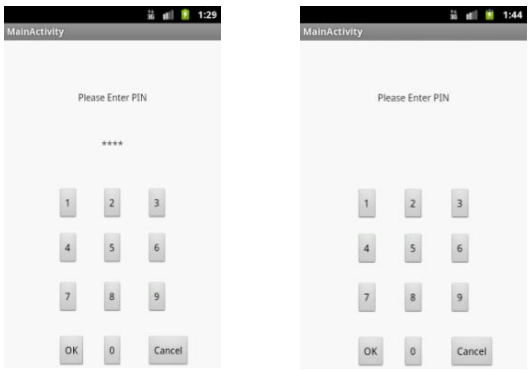
Neural networks and K-means clustering are again used but here in smart phones authentication.

In the first investigation, fifteen samples were used for training and the remaining five for testing. At first, a reference template was constructed for each user containing the hold times of entering the four keys and the latencies periods between the keys. Neural networks were applied on hold times and latencies, the feature vectors were classified using feedforward network.

In the second investigation, fifteen samples were used for training and the remaining five for testing. At first, a reference template was constructed for each user containing the finger touch pressures and finger contact sizes. Neural networks were applied on finger pressures and contact sizes, the feature vectors were classified using feedforward network.

The data was acquired by Android functions, the pressure value was obtained via the getPressure() method while the touch area was captured through the getSize() method. The pressure exerted on the device was obtained in kilopascals while the getsize() functions returns the size of the current contact area. This method returns the size in pixels corresponding to the area touched by the finger [22]. Both functions return a value ranging from 0 to 1 so there is no need for normalization.

Figure 2: Snapshots of Mobile APPLICATION



3) Experimental Results on Key Hold Times and Latencies

In the verification phase, we divided the participants into 2 groups: a group of 10 representing authorized users and the remaining 10 representing unauthorized users. The performance of biometric systems is usually evaluated by two error rates: (FRR) and (FAR).

Hence, FAR and FRR were calculated for each user, then an average was computed. The experiments produced relatively good results; FRR was 19% and FAR was 27%. A snapshot of the mobile application is shown in Figure 2, it shows the layout of buttons the user will use to enter the PIN.

4) Experimental Results on Finger Touch Pressure and Finger Contact Size

Again neural networks were used as in previous experiments resulting in 83% accuracy. FRR and FAR are then measured for each user then an average was computed. The FRR rate was 12% and FAR was 18%.

It has been shown that finger pressure and finger contact areas acted as distinguishable characteristics among users and provided better results than the traditional keystroke features used in computer authentication. Again as presented in the desktop experiment, k-means clustering was implemented on finger pressure and finger contact area data and resulted in 64% accuracy. To conclude, clustering had a better performance in desktop platform rather than neural networks while the opposite occurred in mobile platform.

VII. CONCLUSION

The investigations have shown that it is feasible to authenticate users based on behavioral biometrics. This study has demonstrated the ability of neural networks and k-means clustering to differentiate between computer users based on keyboard and mouse biometrics with a relatively good degree of accuracy. Each technique was applied on keyboard and mouse biometrics each separately, and then fusion of both kinds of data was implemented on the feature level. Different trials were conducted on a number of users and it has been shown that fusion of keyboard and mouse data produced the best results. Also, an authentication scheme for mobile users based on finger touch area and contact finger area was proposed. Before applying the proposed scheme, experiments were done on key hold times and latencies, which are the most commonly used features in keystroke authentication systems. After various experiments, it has been shown that finger pressure and contact size can act as unique features and resulted in better accuracy than the classical keystroke features applied in desktop authentication. This is can be due to that finger pressure and contact size are considered distinguishable among users using touchscreens rather than holding time and latency.

Keystroke analysis has proven to be a promising technique having achieved good results in both desktop and mobile platforms.

VIII. FUTUREWORK

For future work, we plan to explore other machine learning techniques to have a comparative study on different techniques. Also, a comparative research for various smart phones can be implemented as there is a massive evolving variety in touchscreens technology.

Experiments can be conducted on smart phones with different screen sizes to investigate wether screen size can influence finger touch actions which as a result, can affect the authentication accuracy. Also, investigations can be done on smart phones with stylus pens to examine the distinguishing features for those kinds of smart phones and discover if touch pressure and touch contact size can act as unique features or not.

ACKNOWLEDGMENT

We would like to thank Professor Walter Beagley, Professor of Psychology at Alma College for letting us use the Eye Lines Program to collect the data. Also, we gratefully appreciate Romain Giot, Mohamad El-Abed and Christophe Rosenberger for making their GREYCKeystroke software publicly available to help researchers to create keystroke

dynamics databases. Moreover, special thanks to all colleagues who devoted their time to assist us in data gathering.

REFERENCES

[1] J.M. Kizza, Ethical and Social Issues in the Information Age, Texts in Computer Science, Springer- Verlag, 2010.

[2] John Chirillo and Scott Blaul, Implementing Biometric Security, Wiley Pub.,2003, Vol. 14, 9780764525025.

[3] Dmitry O. Gorodnichy, Evolution and evaluation of biometric systems, IEEE Symposium on Computation Intelligence for Security and Defense Applications, 2009.

[4] Prof. V. M. Mane and Prof. (Dr.) D. V. Jadhav, Review of Multimodal Biometrics: Applications, challenges and Research Areas, International Journal of Biometrics and Bioinformatics (IJBB), Vol. 3, Issue 5, 2009.

[5] Jos Alberto Hernndez-Aguilar, Crispin Zavala, Ocotln Daz, Gennadiy Burlak, Alberto Ochoa and Julio Csar Poncem, Biometric Data Mining Applied to On-line Recognition Systems, InTech, 2011.

[6] Romain Giot, Mohamad El-Abed and Christophe Rosenberger, GREYC Keystroke: a Benchmark for Keystroke Dynamics Biometric Systems, IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2009.

[7] Edmond Lau, Xia Liu, Chen Xiao and Xiao Yu, Enhanced User Authentication Through Keystroke Biometrics, Massachusetts Institute of Technology, 2004.

[8] Deian Stefan, Xiaokui Shu and Danfeng (Daphne) Yao, Robustness of keystroke dynamics based biometrics against synthetic forgeries, Journal of Computers and Security 3 I, 2011.

[9] Kenneth Revett, Sergio Tenreiro de Magalhaes and Henrique Santos, DataMining a Keystroke Dynamics Based Biometrics Database Using Rough Sets, IEEE, 2005.

[10] Shivani Hashia, Chris Pollett and Mark Stamp, On Using Mouse Movements as a Biometric, San Jose State University, 2008.

[11] AdamWeiss, Anil Ramapanicker, Pranav Shah, Shinese Noble and Larry Immohr, Mouse Movements Biometric Identification: A Feasibility Study, Seidenberg School of CSIS, Pace University, 2007.

[12] Seong-seob Hwang, Sungzoon Cho, and Sunghoon Park, Keystroke dynamics-based authentication for mobile Devices, Journal of Computers and Security, 2009.

[13] N.L. Clarke, S.M. Furnell, Advanced user authentication for mobile devices, Journal of Computers and Security, 2007.

[14] Julio Angulo and Erik W?stlund, Exploring Touchscreen Biometrics for User Identification on Smart Phones, Karlstad University, 2011.

[15] Saira Zahid, Muhammad Shahzad, Syed Ali Khayam and Muddassar Farooq, Keystroke-based User Identification on Smart Phones, Springer Verlag, 2009.

[16] Michal Choras and Piotr Mroczkowski, Keystroke Dynamics for Biometrics Identification, Springer- Verlag, 2007.

[17] S.Benson Edwin Raj and A. Thomson santhosh, A Behavioral Biometric Approach Based on Standardized Resolution in Mouse Dynamics, IJCSNS International Journal of Computer Science and Network Security, Vol. 9, No. 4, 2009.

[18] http://www.alma.edu/el/, [accessed 10/09/2011].

[19] Meier, R., Professional Android 2 Application Development, John Wiley and Sons Inc., 2010.

[20] Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques, Elsevier, 2006, Second Edition, 81-312-0535-5.

[21] Arun Ross and Rohin Govindarajan, Feature Level Fusion Using Hand and Face Biometrics, SPIE Conference on Biometric Technology for Human Identifiication II, Vol. 5779, pp. 196-204, 2005.

[22] Ting-Yi Chang, Cheng-Jung Tsai, Jyun-Hao Lin, A graphical-based password keystroke dynamic authentication system for touch screen handheld mobile devices, The Journal of Systems and Software, 2012.

Fusion of Biogeography based optimization and Artificial bee colony for identification of Natural Terrain Features

Priya Arora
Mtech student, CSE
RIEIT, PTU
Punjab, India

Harish Kundra
HOD, CSE
RIEIT, PTU
Punjab, India

Dr. V.K Panchal
Associa te Director, Sci 'G'
DTRL, DRDO
Delhi, India

***Abstract*— Swarm Intelligence techniques expedite the configuration and collimation of the remarkable ability of group members to reason and learn in an environment of contingency and corrigendum from their peers by sharing information. This paper introduces a novel approach of fusion of two intelligent techniques generally to augment the performance of a single intelligent technique by means of information sharing. Biogeography-based optimization (BBO) is a recently developed heuristic algorithm, which proves to be a strong entrant in swarm intelligence with the encouraging and consistent performance. But, as BBO lacks inbuilt property of clustering, its behavior can be replaced with the honey bees of artificial bee colony (ABC), a new swarm intelligent technique. These two methods can be combined to create a new method which is easy to implement and gives more optimized results than the results when BBO is used. We have successfully applied this fusion of techniques for classifying diversified land cover areas in a multispectral remote sensing satellite image. The results illustrate that the proposed approach is very efficient than BBO and highly accurate land cover features can be extracted by using this approach.**

Keywords- Biogeography-based Optimization; Artificial bee colony; Hybrid swarm intelligence; Image classification; Multi spectral dataset.

I. INTRODUCTION

Swarm Intelligence is an venture to design algorithms or distributed problem solving methods stimulated by collective behavior of social insect colonies. Examples of systems studied by Swarm Intelligence are colonies of ants and termites, schools of fish, flock of birds, herds of land animals. Various Swarm based approaches have been developed till now. A novel approach added recently in this category called Biogeography-based Optimization proposed by Dan Simon (2008) [1]. Karaboga and Basturk (2005), proposed artificial bee colony algorithm for solving many optimization problems which is easy to implement and robust. The key to sustaining global, self- organized behavior is social interaction.

The fundamental principle of these techniques is cooperation and sharing of knowledge. The basic of the increased intelligence is the shared information discovered individually and communicated to the swarm by different mechanisms of social interactions. In this paper, a novel integrated SI algorithm is developed by synthesizing Biogeography based optimization [2] and artificial bee colony optimization [3]. We have tried to make the BBO technique more intelligent simply by making the species in BBO to work according to the honey bees of ABC. In this way, intelligent solutions to problems naturally will emerge from the self-organization and communication among simple individuals. To validate the effectiveness of the proposed approach, it is applied for the classification of various natural terrain features which itself is an optimization problem. We have compared our results with BBO to show the performance of our proposed technique.

In this paper, the first section is an introduction to swarm intelligence. The second section gives detail of the BBO and ABC. The third section discusses the proposed approach i.e. Fusion of BBO and ABC. The fourth section discusses the implementation of the proposed concept for classifying natural terrain features. The last section presents the concluding remarks.

II. BACKGROUND

A. Biogeography Based Optimization

The idea of Biogeography Based Optimization (BBO) was first presented in December 2008 by D. Simon[2]. Biogeography is the study of the geographical distribution of biological organisms. The mindset of the engineer is that we can learn from nature. This motivates the application of biogeography to optimization problems. Just as the mathematics of biological genetics inspired the development of genetic algorithms (GAs), and the mathematics of biological neurons inspired the development of artificial neural networks, mathematics of biogeography as the basis for the development of a new field: biogeography-based optimization (BBO).

Mathematical models of biogeography describe how species migrate from one island to another, how new species arise, and how species become extinct. The species are migrated to the suitable habitats i.e, feature islands. These habitats are actually the decisions, for the resultant groups. Various methods decide which species are moved to which habitat using a fitness function [2]. The fitness function is actually the information that is shared among all the habitats in order to decide the suitable habitat for migration of each species. Geographical areas that are well suitable as residences

for the biological species are said to have a high suitability index (HSI) [2]. Features that correlate with HSI include factors such as rainfall, diversity of vegetation, land area and temperature. The variables characterize habitability are called suitability index Variables (SIVs). SIVs can be considered the independent variables of the habitat, and HSI can be considered as the dependent variable.

Habitats with a high HSI have many species that emigrate to nearby habitats, simply by the virtue of the large number of species that they host. Habitats with a high HSI have a low species immigration rate because they are already nearly saturated with species. Habitats with a low HSI have a high species immigration rate because of their sparse populations. Biogeography is nature's way of distributing species and is analogous to general problem solutions. Suppose that we are presented with a problem and some candidate solutions. A good solution is analogous to an island with a high HSI and a poor solution rep resent an island with a low HSI. We call this approach to problem solving Biogeography Based Optimization [1].

B. Artificial Bee Colony

ABC was originally proposed by Dervis Karaboga [3] under the inspiration of collective behavior of honey bees. Real bee colony is one of the natural societies with most specialized social behaviors. ABC algorithm consists of three kinds of bees: employed bees, onlooker bees, and the scout bees[5]. The goal of the whole colony is to maximize the amount of nectar.

1. Employed bees determine a food source considering the surrounding food source in their memories.
2. Employed bees inform the onlooker bees in the hive about the food source and onlooker bees select a food source.
3. Employed bee whose source is abandoned becomes a scout bee and start to search for a new food source.

All the information about the currently rich food source are available on the dance floor area and the onlooker watches numerous dances performed by the employed bees and chooses the profitable food source. The onlooker bee decides the profit using the probability values of the food sources. The recruitment is thus proportional to the profitability of a food source. Basically, there are two important functions that supports the algorithm[5].

$$P_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_i} \quad \text{...(1)}$$

$$\mathbf{V_{ij}} = \mathbf{X_{ij}} + \mathbf{\Phi_{ij}}(\mathbf{X_{ij}} - \mathbf{X_{kj}}) \quad \text{...(2)}$$

where P_i is the probability value associated with $\mathbf{i_{th}}$ food source that calculated by the Eq a. An onlooker bee selects a food source relying on P_i. In this equation, $\boldsymbol{fit_i}$ represents $\mathbf{i_{th}}$ food sources nectar amounts, which is measured by employed bees and SN is the number of food source which is equal to the number of employed bees. Fitness is calculated using following equation;

$$fitness(i) = \begin{bmatrix} 1/(1+fnc_i), & fnc_i \geq 0 \\ 1+abs(fnc_i), & fnc_i < 0 \end{bmatrix}$$

the $\boldsymbol{fnc_i}$ in the equation is the cost function of the quality of source. Greedy selection is applied to select the best source. In the real world problems, $\mathbf{X_{ij}}$ and $\mathbf{X_{kj}}$ represents the different old food source positions. The difference between these two positions is the distance from one food source to the other one. D is number of optimization parameters. φ_{ij} is a random number between [-1,1] and controls the distance of a neighbor food source position around $\mathbf{X_{ij}}$[5].

III. Integrated Algorithm Based On BBO And ABC

BBO and ABC are two outstanding swarm intelligence methods. Biogeography describes how various species migrate from one island to another according to their suitability conditions. The heuristic method decides which species are moved to which habitat using a fitness function. The fitness function is actually the information that is shared among all the habitats in order to decide the suitable habitat for migration of each species. In ABC the bees distributes themselves to various food locations and finally on finding most suitable food source the bees are followed by rest of their bee mates to that particular food source. In our proposed technique we have replaced the behavior of species in BBO by the behavior of honey bees of ABC.

The main idea of this proposed approach is to enhance the quality of BBO by bartering information from the other intelligent technique i.e. ABC. Thus the species will be randomly distributed to the habitats rather than moving with the whole group of species. All the featured habitats will be explored by the species. On finding a habitat whose HSI is near about the defined fitness value, neighborhood of that specific habitat will be further explored by the species in order to find a better place to migrate. Thus on finding a habitat whose HSI is high, the species will exchange the information of that habitat with the rest of the species waiting for them (Universal habitat) as the employed bees exchange information related to different food sources with the onlooker bees. Now the whole group will collectively migrate to that suitable habitat for their migration.

In this way we can create knowledge by sharing information between two intelligent techniques to make a technique more intelligent. Hence this proposed intelligent computing hybrid approach would be able to solve complex problems providing more efficient results which would be better than the results of BBO. The framework of the algorithm could be described as follows :

1. Initialize solutions belongs to different habitats.

2. Select species from the universal habitat and randomly distribute the species to the feature habitats (It will be executed according to the employed bees' search)

3. Calculate the HSI (fitness value) for each habitat.

4. If the HSI calculated is within the threshold value then explore the neighborhood solution else go to step 2.

5. Select the most suitable habitat for migration.

6. End.

In order to validate this proposed scheme, it will be experimented on a set of data.

IV. EXPERIMENT SETUP

We have applied this fusion of BBO and ACO on the remote sensing multi- spectral, multi resolution and multi sensor image of Alwar area in Rajasthan with dimensions 472 * 572 for classifying the various terrain features. The area is selected since it contains a good variety of land use features like urban, water body, rocky, vegetation and barren areas. The 4 spectral bands are in the visible bands namely : red, green, near - infrared (NIR), and middle infra - red (MIR) from LISS- 111 sensor of Indian Remote Sensing sat satellite Resource sat -1. RS1 and RS2 are the bands of the same area taken from Canadian Radarsat-1 satellite. The seventh image is digital elevation model (DEM) of the area. The ground resolution of these images from LISS-111 and Radarsat-1 images are 23.5 m and 10 m respectively. The DEM dataset is also generated from SAR interferometer using RS1 and RS2 and have 25 -meter resolution and is taken from LISS=11, sensor and is provided to us courtesy of DTRL(Remote Sensing) Lab, DRDO, Delhi. The 7 band Satellite image of Alwar Area in Rajasthan is given in figure 1.

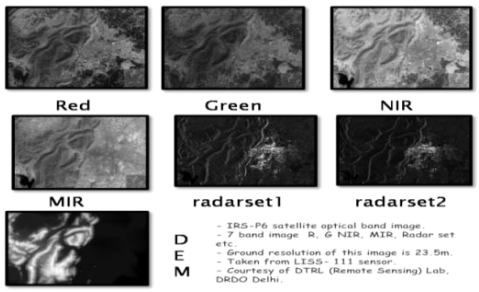


Figure1. 7-Band Satellite Image of Alwar region, Rajasthan.

A. *Proposed Algorithm for classification*

Assumptions

Food sources are the land cover features- water, vegetation, urban, rocky, and barren.

Input :Dataset of Multi- spectral satellite image.Output : Classified image**.**

Step 1: Initialization

- Consider original image as species of universal habitat.
- Consider each feature class as one habitat. Total no of habitat = Universal + Feature habitat.

Step 2: Allocation of species according to employed bees.

- Distribute species to various habitats (according to employed bees' search)

Step 3 : Calculate HSI of each habitat.

- Evaluate HSI of each habitat here calculated using Mean of Standard Deviation as

$$HSI = \sum_{k=1}^{mb} S(i,k)/mb$$

where i is for each habitat(different feature exist on the study area), varies from 1 to mb.

S_{ik} represents **Standard Deviation** of *k*th band of *i*th habitat (k varies for each availasble band, i.e. 1 to mb, where mb is the number of band available)

$$S_{i,k} = \frac{\sqrt{\sum_{j=1}^{n_i} V_i(k,j) - M(i,k)}}{n_i}$$

where

- $\mathbf{n_i}$ is the training data available for each land cover type.
- $V_i(k,j)$ represents the pixel value of ***k*th** band in ***jth*** training pixel of ***i*th** habitat.
- $M_{i,k}$ is the mean of ***k***th band in the ***i***th habitat

Step 4*:* Selection of Habitat

- **If** calculated HSI is within threshold then
  - Continue searching the neighborhood of that habitat for better solutions.
  - Absorb species to that habitat.

- **else**
  - go to step 2.

- Check for the other habitats and recalculate the HSI.

- End loop when all pixels classified**.**

- Display the output classified image.

B. *Flowchart*

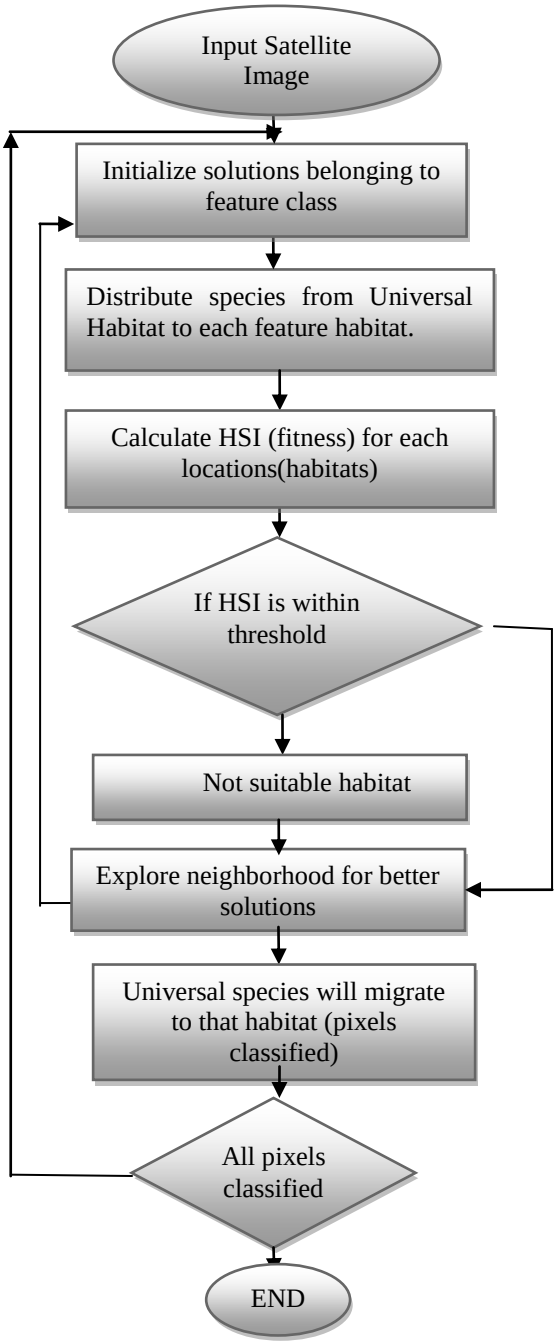


Figure3 shows the satellite false color image and Figure 4 is the classified image of Alwar region after applying BBO-ABC.

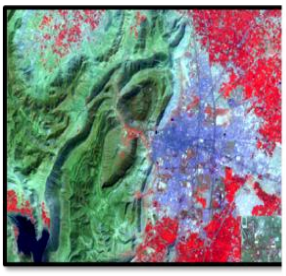


FIGURE 3. ORIGINAL IMAGE

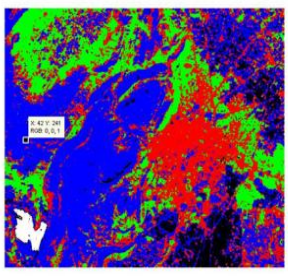


FIGURE 4. CLASSIFIED IMAGE TABLE TYPE

C. Accuracy statement

A classification is not complete until its accuracy is assessed [6]. Practically it is not feasible to test every pixel of a class known. Instead a set of reference pixels is used. These are the points on the classified image for which actual features are known. The reference pixels are randomly selected. The main goal of accuracy assessment is to quantitatively determine how effectively pixelare grouped into the correct feature classes in the area under investigation. In our proposed work the classification accuracy is expressed using classification error matrix. For the validation process we have taken into consideration following number of pixels;

Table1. Error Matrix when BBO -ABC is applied

Feature	Water	Veg	Urban	Rocky	Barren	Total
Water	68	0	0	0	0	68
Veg	0	109	1	0	0	110
Urban	0	0	112	0	4	116
Rocky	0	0	0	96	0	96
Barren	0	0	26	0	59	85
Total	68	109	139	96	63	475

- 68 water pixels.
- 109 vegetation pixels.
- 139 urban pixels.
- 96 rocky pixels.
- 63 barren pixels.

The Kappa coefficient of the Alwar image can be calculated by applying following formula to the Error Matrix;

$$\hat{k} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{i+} \cdot x_{+i})}$$

- r = number of rows in the error matrix (r= 5 in our case)
- x_{ii} = the number of observations in row i and column i (on the major diagonal)
- x_{i+} = total no of observation in row i (shown as marginal total to right of the matrix)
- x_{+i} = total of observations in column i (shown as marginal total at bottom of matrix)
- N = total number of observations included in matrix.(N= 475 in our case)

The Kappa coefficient of the Alwar image is 0.917, which indicates that an observed classification is 91.7 percent better than one resulting from chance whereas the value of Kappa coefficient in case of BBO is 0.69121 [4]. So it is clearly understood that our classifier has good classification accuracy as compared to BBO.

V. Conclusion And Future Work

This paper introduces the fusion of two swarm intelligent techniques BBO and ABC. The particles in a swarm try to solve a problem as a group by using the information contained by their peer mates. This sharing of information enables swarm particles to be more efficient and to achieve goals that they could not achieve individually. To check the efficiency of our proposed approach it has been used for the classification of natural terrain features. Experimentally we have proved that the proposed algorithm yields an approximate but usually good solution to the classification problem for satellite images. The kappa coefficient of 0.917 also supports our algorithm efficiency from our table which are better than the results when BBO whose kappa coefficient is 0.69121. We can say that for the given dataset we have been able to classify water, vegetation, rocky areas perfectly. Therefore we illustrate the information sharing concept in the swarm intelligence techniques by means of a terrain understanding application. In future, the concept of information sharing can be exploited further to adapt to other applications such as groundwater exploration through case based reasoning using any of the above SI techniques. This practice can be easily being extended for other global optimization problems.

References

[1] Ralph W. Kiefer, Thomes M. Lillesand, "Principles of Remote Sensing",2006.

[2] Dan Simon, "Biogeography Based Optimization", IEEE transactions on evolutionary computation, vol. 12, no. 6, December 2008

[3] D Karaboga , " An idea based on honey bee swarm for numerical optimization TR-06", 2005.

[4] V.K Panchal, D Bhugra , S Goel, V Singhania, " Study On The Behaviour Of BBO Over Natural Terrain Features", Electronics Computer Technology (ICECT), 3rd International Conference,2011.

[5] Yuk Ying Chung, Wei-Chang Yeh, Noorhaniza Wahid, Ahmad Mujahid, Ahmad Zaidi, "Artificial Bee colony based data mining Algorithms for classification task",vol 5, no. 6; August 2011.

[6] T.MLillesand, R.W Keifer, J.W Chipman, "Remote Sensing and Interpretation", New York: John Wiley & Sons Ltd, pp, 586 -592(2004).

Authors Profile

Priya Arora has done B-Tech (Hons) in Information Technology & scored 78 % marks from Punjab Technical University, Jalandhar (India) in 2010 and Mtech in CSE from Punjab Technical University with 70% marks.

Harish Kundra is an Assistant Professor and Head of Department in CSE & IT Department of Rayat Institute of Engineering &IT, Railmajra, Punjab, India. He has presented 15 papers in International/National journals. He is a lifetime member of Indian Society of Technical Education.

V.K Panchal is Associate Director at Defence Terrain Research Lab, New Delhi, Associate Member of IEEE, (Computer Society) and Life Member of Indian Society of Remote Sensing. Chaired sessions & delivered invited talks at many national & international conferences. Research interest are in synthesis of terrain understanding model based on incomplete information set using bio-inspired intelligence and remote sensing.

High Performance Speed Sensorless Control of Three-Phase Induction Motor Based on Cloud Computing

Z.M. Salem
Electrical Engineering Department
Faculty of Engineering
KFS University, Egypt

M.A.Abbas
Electrical Engineering Department
Faculty of Engineering
King Khaled University, KSA

***Abstract*— Induction motor is a cast of alternating current motor where charge endures allotted to the rotor close-at-hand deputation of conductive charge. These motors are broadly applied in industrial claim due to they are arduous along with adhere no contacts. The speed controller of deltoid phase induction motor is applied to alleviate the aberration of speed. The central constructivist of this paper is to accrue the performance of speed sensorless control of three phase induction motor. To increase its performance, this paper presents a modified method for speed controller of an indirect vector-controlled induction motor drive using cloud computing technique. Our methodology depends on speed sensorless scheme to obtain the speed signal feedback; the speed estimator is based on model reference adaptive control that uses the stator current and rotor flux as state variables for estimating the speed. In this method, the stator current error is represented as a function of first degree of the estimated speed error. An analysis and simulation of the tried algorithm is birthed and applied easing a TMS320C31 floating-point notational alert Processor. And accumulate the action of the three phase induction motor we conceived our appraisals affixed to the accountant based on cloud computing tactics. This intelligent policy uses the guidelines of the speed controller efficiently. Simulation and experimental results depicted that the motor speed is decelerated articulately to destine its illusion apprise without above and inferior smack and with about zero steady state error. The apprised accelerate alert and its dispatching buoy amassed off line from burlesque. After effects display an advantageous affinity among the accounted speed alert and it's dispatching allocated as well as aped speed flares.**

Keywords- cInduction motor; Cloud computing control; Sensorless control; Vector control; Observers; Modeling; Identification.

I. INTRODUCTION

In the last few decades, induction motor (IM) particularly squirrel-cage, has been recognized as a workhorse in the industry because it have many inherent advantages like simplicity, reliability, low cost and virtually maintenance-free. Integrating this motor type with a reliable controller became quite important for numerous industrial applications. Currently, indirect field oriented control technique is one of the first choice controllers for high-performance induction motor drives, however, rotor speed or position feedback data is essential for proper operation [1-4].Tachogenerators or optical shaft encoders can be used for this purpose; however, besides the high cost, these direct speedsensors often spoil the ruggedness, reliability, and simplicity of an induction motor drive. Moreover, such sensitive devices require careful mounting and alignment, and need special attention to be paid to electrical noise interference with their output data. Furthermore, an exact servo control performance is sometimes required in an operating environment where the attachment of a direct speed sensor is impossible. To reduce total hardware complexity and cost and to increase mechanical robustness, it becomes advantageous to replace these direct sensors by some other speed estimation algorithm preserving the high system performance [5-8].

Various control algorithms have been proposed for the speed-sensorless control of an induction motor [9-11]. These sensorless algorithms are mainly based on an estimated flux and speed feedback signals. Speed observer based on the theory of model reference adaptive system (MRAS) is one of the most popular techniques that is usually implemented for speed sensorless induction motor drives [12-14]. In this algorithm, the rotor flux is estimated from the stator equation (considered to be the reference model, once it does not depend on the speed) and also using the rotor equation (adaptive model). The speed is then obtained by the use of an adaptive law having the cross product of the two estimated signals as inputs [7]. Rotor flux or stator back emf may be used to make a reference function, and then the motor speed is estimated using MRAS. Conventionally, PI controllers are usually used for implementing MRAS flux and speed observers. In spite of its simplicity, the performance of a PI-based observer is often deteriorated due to system nonlinearity originated from its parameter uncertainty and mismatch [5, 6].

Cloud computing is the ease of accounting assets that are allotted as an agency above a crossway. The appoint accesses from the use of a cloud-shaped alert as an emptiness for the abstruse infrastructure it accommodates in algebra blueprints. Cloud computing accredits distant aids with a consumer's attestation, software additionally appraisal. As cash registers benefited accrual accepted, scientists and technologists canvassed channels to construct large-scale apprising activity achievable to additional consumers accomplished era allocating, analyzing with algorithms to ascribe the best ease of the infrastructure, platform and addresses with ranked access to the adding machine and advantageousness for the

back consumers [15-18]. John McCarthy brainstormed in the 1960s that "appraisal may someday be adjusted as a communal application." Almost accomplished the modern-day cachets of cloud computing, the comparison to the anode activity and the ease of communal, confidential, authority, and citizenry buds, were collectively canvassed in Douglas Parkhill's 1966 brochure, The confront of the Computer Utility. Other scholars have shown that cloud computing roots go all the way back to the 1950s when scientist Herb Grosch postulated that the entire world would operate on speechless accomplishments began adjacent about 15 ample attestation centers[19-21]. Due to the cost of these arrogant calculators, common concerns additionally irrelevance entities could endowment themselves attendant adding ability accomplished age allocating and numerous agencies, akin as GE's GEISCO, IBM subsidiary the agency buffet activity.

The aim of this paper is to design and implement a speed control scheme of 3-phase induction motor drive system using PI based cloud computing, in which, the system control parameters are adjusted by cloud computing based system. The foremost compensation of cloud computing over the conformist controllers are that The cooperative accessibility analogously impressive cerebration networks, despicable quantity abacuses along with store apparatuses as well as the widespread adoption of hardware virtualization, service-oriented anatomy, autonomic, as well as employment accounting acquires administered to an ample amplification[22-24]. Also, a stator current based MRAS for speed estimation of a sensorless induction motor drive is presented. The speed estimation error is continuously reduced to zero using a PI controller as an adaptive low. The effectiveness of the proposed method is tested at different operating conditions. A floating-point Digital Signal Processor (DSP) TMS320C31 control board with a hardware/software interface has been used to implement the proposed method for speed estimation. Simulation and experimental results are presented and discussed

II. System Description

The proposed system intended for performance analysis of a cloud computing based model reference adaptive system (MRAS) speed estimator of an indirect vector controlled induction motor drive is shown in Fig. 1. Speed feedback signal is obtained by a MRAS speed estimation block instead of direct measurement via a shaft encoder. Voltage and current signals are obtained by Hall-effect sensors and send to the DSP via its A/D input ports. This speed estimator uses the accessible terminal signals representing stator phase currents and voltages as an input data after being manipulated by suitable axes transformations. The field oriented control (FOC) block receives the torque command T^* obtained from the speed controller while the flux command λ_{dr}^* is maintained constant. The FOC block performs the slip calculation and generates the current command components i_{qs}^e and i_{ds}^e in a rotating reference frame. These components are further manipulated by axes transformations to obtain the abc current command components i_a, i_b, and i_c.

The axes transformations used for the present system are expressed as follows;

$$\begin{bmatrix} i_{qs}^{s*} \\ i_{ds}^{s*} \end{bmatrix} = \begin{pmatrix} \cos\theta_s & \sin\theta_s \\ \sin\theta_s & \cos\theta_s \end{pmatrix} * \begin{bmatrix} i_{qs}^{e*} \\ i_{ds}^{e*} \end{bmatrix}$$

where θ_s represents the sum of the slip and rotor angles.

$$qds \rightarrow abc \begin{Bmatrix} i_{as}^{s*} = i_{qs}^{s*} \\ i_{bs}^{s*} = -\frac{1}{2} i_{qs}^{s*} - \frac{\sqrt{3}}{2} i_{ds}^{s*} \\ i_{cs}^{s*} = -\frac{1}{2} i_{qs}^{s*} + \frac{\sqrt{3}}{2} i_{ds}^{s*} \end{Bmatrix}$$

III. Mathematical Model

This section presents the mathematical model of the induction motorto revise the recital of the scheme at diverse working settings. In addition a detailed analysis of a rotor speed estimator and the main concept of PI controller using cloud computing strategy.

A. Induction Motor Model

Squirrel-cage induction motor is represented in its d-q dynamic model. This model represented in synchronous reference frame is expressed as follows;

$$\begin{bmatrix} V_{qse}^e \\ V_{dse}^e \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} R_s + pL_\sigma & \omega_e L_\sigma & p\frac{L_m}{L_r} & \omega_e \frac{L_m}{L_r} \\ -\omega_e L_\sigma & R_s + pL_\sigma & -\omega_e \frac{L_m}{L_r} & p\frac{L_m}{L_r} \\ -R_r L_m & 0 & R_r + pL_\sigma & (\omega_e - \omega_r)L_m \\ 0 & -R_r L_m & -(\omega_e - \omega_r)L_m & R_r + pL_\sigma \end{bmatrix} \begin{bmatrix} I_{qs}^e \\ I_{ds}^e \\ \lambda_{qr}^e \\ \lambda_{dr}^e \end{bmatrix} \quad (1)$$

The electromechanical equation is also given by;

$$T_e - T_L = J\frac{d\omega_r}{dt} + B\omega_r \quad (2)$$

Where, the electromagnetic torque is expressed as;

$$T_e = \frac{3}{2}\frac{p}{2}.\frac{L_m}{L_r}(I_{qs}^e \lambda_{dr}^e - I_{ds}^e \lambda_{qr}^e) \quad (3)$$

Assuming the stator applied voltage V is known, and the stator current can be obtained directly via measurements, the flux vector can be obtained by integration as follows

$$\lambda_s = \int (V - R_s I_s)dt \quad (4)$$

This equation is often called the stator flux observer.

B. Speed Estimator

A rotor speed estimator is used to study the performance of the system at different operating conditions. The estimator depends on both MRAS and adaptive speed observer which are based on rotor flux. Speed estimation procedure of the proposed method is illustrated by the following analysis.

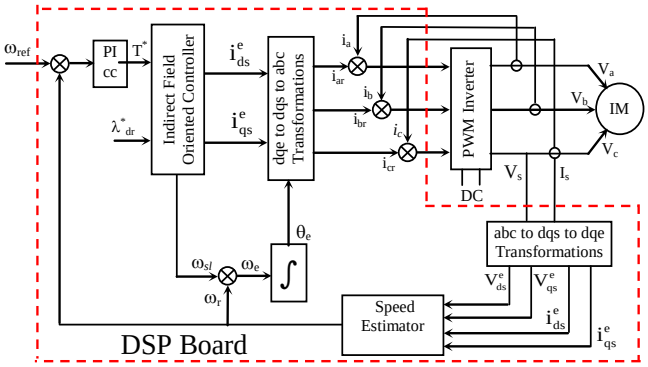


Fig. 1 Block Diagram of the Proposed Speed Sensorless Control System

The stator current is represented as:

$$i_{ds} = \frac{1}{L_m}\left[\lambda_{dr} + \omega_r T_r \lambda_{qr} + T_r p\lambda_{dr}\right]$$
$$i_{qs} = \frac{1}{L_m}\left[\lambda_{qr} - \omega_r T_r \lambda_{dr} + T_r p\lambda_{qr}\right] \quad (5)$$

Using the above Eqns, the stator current is estimated as

$$\hat{i}_{ds} = \frac{1}{L_m}\left[\lambda_{dr} + \hat{\omega}_r T_r \lambda_{qr} + T_r p\lambda_{dr}\right]$$
$$\hat{i}_{qs} = \frac{1}{L_m}\left[\lambda_{qr} - \hat{\omega}_r T_r \lambda_{dr} + T_r p\lambda_{qr}\right] \quad (6)$$

The difference in the stator current is obtained as

$$i_{ds} - \hat{i}_{ds} = \frac{T_r}{L_m}\lambda_{qr}\left[\omega_r - \hat{\omega}_r\right]$$
$$\hat{i}_{qs} - i_{qs} = \frac{T_r}{L_m}\lambda_{dr}\left[\omega_r - \hat{\omega}_r\right] \quad (7)$$

Equation (7) may be rewritten as:

$$(i_{ds} - \hat{i}_{ds})\lambda_{qr} = \frac{T_r}{L_m}\hat{\lambda}_{qr}^2\left[\omega_r - \hat{\omega}_r\right]$$
$$(\hat{i}_{qs} - i_{qs})\lambda_{dr} = \frac{T_r}{L_m}\hat{\lambda}_{dr}^2\left[\omega_r - \hat{\omega}_r\right] \quad (8)$$

Since the stator current error is represented as a function of estimated speed, an adaptive flux observer can be constructed from the machine model equation. The model outputs are the estimated values of the stator current vector $\hat{i}_s$ and the rotor flux linkage vector $\hat{\lambda}_r$.

From Eqn. (8),

$$(i_{ds} - \hat{i}_{ds})\lambda_{qr} + (\hat{i}_{qs} - i_{qs})\lambda_{dr} = \frac{T_r}{L_m}(\hat{\lambda}_{qr}^2 + \hat{\lambda}_{dr}^2)\left[\omega_r - \hat{\omega}_r\right] \quad (9)$$

Hence, the error of the rotor speed is obtained as follows:

$$\omega_r - \hat{\omega}_r = [(i_{ds} - \hat{i}_{ds})\lambda_{qr} - (i_{qs} - \hat{i}_{qs})\lambda_{dr}]/K$$
$$\text{where} \quad K = \frac{T_r}{L_m}(\lambda_{qr}^2 + \lambda_{dr}^2) \quad (10)$$

The right hand term seems as the term of speed calculation from adaptive observer, so the speed can be calculated from the following equation,

$$\hat{\omega}_r = \frac{1}{K}[(K_p(i_{ds} - \hat{i}_{ds})\lambda_{qr} - (i_{qs} - \hat{i}_{qs})\lambda_{dr}) + (K_I\int(i_{ds} - \hat{i}_{ds})\lambda_{qr} - (i_{qs} - \hat{i}_{qs})\lambda_{dr})dt] \quad (11)$$

The speed estimation procedure represented by Eqns 5 to 11 is illustrated by the block diagram shown in Fig. 2.

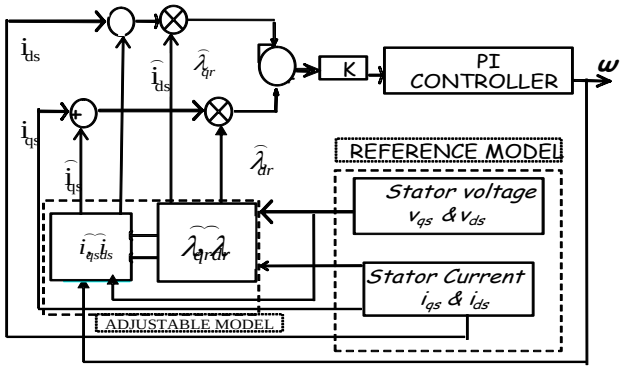


Fig. 2 Block Diagram for Speed Estimation Procedure

C. Cloud Computing Based PI Controller

Cloud computing is changing the whole manufacturing, lofty-concert subtracting and individual statistics allocation and organization. In cloud computing, computing power is abounding as a utility, comparable to electrical energy or irrigate. Overhaul supplier can centrally administer, preserve, and advance working out possessions, divesting the lumber on or after diminutive trade landlords or individuals who do not include the proficiency or funds to lever the speedy-varying subtracting transportation. By means of the cloud for elevated-recital computing can significantly diminish the entirety charge of tenure by eradicating the necessitate to preserve huge-balance equivalent machinery and their vigor-overriding rule and fresh structures. Starting a charge-efficacy perception, there are tradeoffs in requisites of reserve provisioning specified that a objective chore can be parallelized, a frequent holder for throughput-tilting calculation.

Fig 3 depicts the flowchart of proposed cloud computing algorithm. The baseline of the revise presumes that the whole occupation is executed on one a appliance consecutively on the fastest corporal knob. The greatest substantial lump can leave a work part each a subsequent. Since there are g self-regulating profession elements in the intact workload, the baseline pattern receives ga instants to terminate. This configuration munches through W × ga joules for carrying out the full workload, where W represents a corporeal node's power. Thus, the EDP is EDPbase = (W × ga)(ga) = Wg2a2 [25].

An anticipation-based psychoanalysis is used to conclude a cloud model's completing time and force expenditure. A novel allocation utility is worn to stand for the execution occasion of a practical appliance with more than one job element.

When sovereign models are supplementary from a consistent allocation, the outline's allotment purpose is liable to come near an ordinary allotment according to the central limit theorem.9, this theorem proves that when we add more autonomous modesl into the rundown, the précis's division will develop into extra approximating a standard supply. The mean and variance of the normal distribution representing the total execution time of a virtual machine responsible for m/p job units. First, we calculate the mean and variance for the original uniform distribution, U(a,(a + ((b – a)p)/n)):

$$mean = a + \frac{(b-a)p}{2n}$$

$$\text{var}iance = \left(\frac{(b-a)p}{\sqrt{12}n}\right)^2$$

The innermost perimeter theorem depicts the rundown of m/p sovereign trials from this allocation will develop into a regular allotment with the subsequent mean and variance:

$$N\left(\frac{m}{p}\left(a+\frac{(b-a)p}{2n}\right),\left(\sqrt{\frac{m}{p}}\times\frac{(b-a)p}{\sqrt{12}n}\right)^2\right) = N(\mu,\sigma^2)$$

For ease, μ and σ2 is used to signify the distribution's mean and variance. All in all, when using p effective machinery, apiece appliance's implementation time will follow the normal distribution, N(μ, σ2)[25].

IV. Simulation And Experimental Results

The proposed control system represented by Fig. 1 is designed and implemented for a simulation and experimental investigation.

Simulation is carried out using the general purpose simulation package Matlab/Simulink, while experimental study is implemented using a TMS320C31floating-point Digital Signal Processor (DSP) hosted on a personal computer. Simulation and experimental results are presented to show the effectiveness of the proposed drive system based cloud computing based controller instead of PI controller at different operating conditions.

For studying the performances of proposed system, a series of simulations and measurements have been carried out. In this respect, the dynamic response of the proposed speed estimation algorithm is studied under both step up and step down changes in the speed command as follow.

A. A-Speed step down change from (80 to 60 rad/sec.

To study the system response of the control system due to a step changes in the command of speed, the motor is subjected to step decrease in the speed command to evaluate its the performance.

At t=1.1 second the motor speed command is changed from 80 rad/sec to 60 rad/sec. Figure 4 shows the motor speed corresponding to this step down changes. It can be seen that the motor speed is decelerated smoothly to follow its reference value without over and under shoot and with nearly zero steady state error.

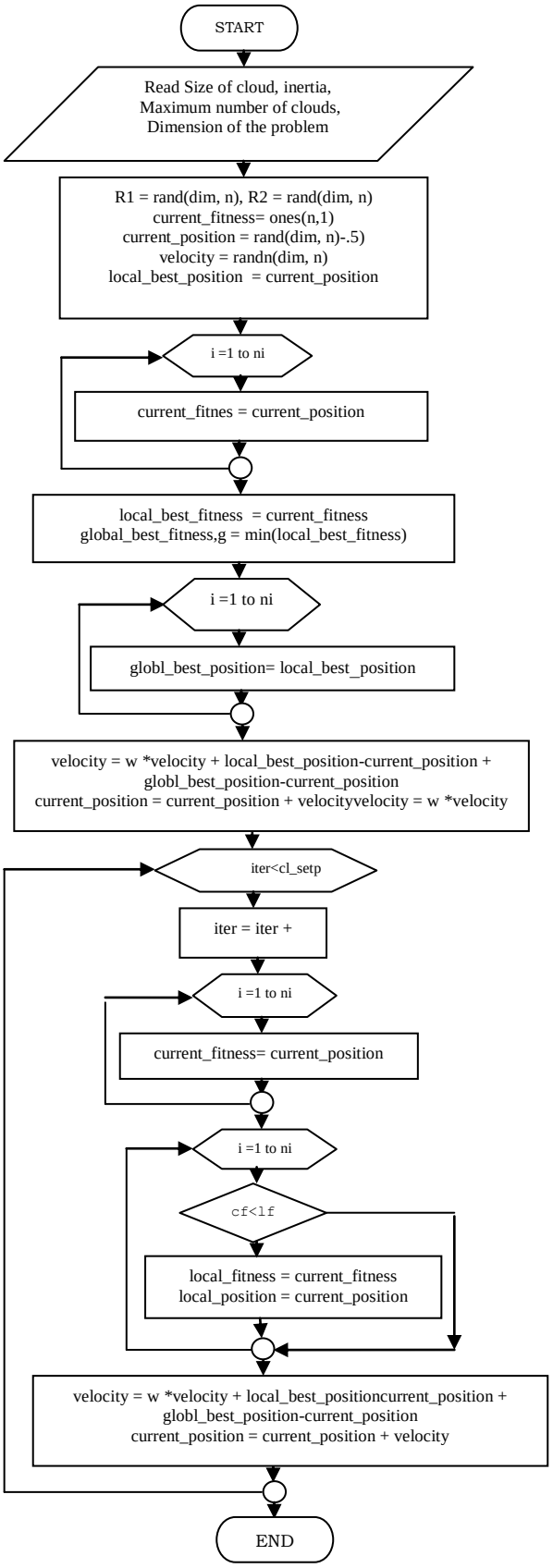


Fig. 3 Flowchart of proposed cloud computing algorithm

Figure 4.a shows the estimated speed signal and its corresponding signal obtained off line from simulation. Figure 4.b shows measured and estimated speed signals obtained in real time. These results show a good correlation between the estimated speed signal and its corresponding measured as well as simulated speed signals. Phase current correspondent to these step varying are depicted in Figs. 5, 6 in that order.

Figures 5a and 5b, symbolizes the phase current and its orientation domination, both from simulation and experimentally. There is a good association among all current signals, whereas figure 6a and 6b shows the three phase currents. These results make certain the efficacy of the projected controller and shows good actions of its self-motivated reaction.

B. B-Speed step up change from (40 to 60 rad/sec)

The second case of dynamic response is due to speed step up change. At t=1.1 second ,The motor is subjected to a command of speed up from 40 rad/sec to -60 rad/sec. Figure 7.a shows the estimated and simulated speed signals obtained off line for this condition. It can be seen that both speed signals have the same profile which are always almost correlated.

Figure 7.b shows the corresponding real time estimated and measured for this condition. Both signals show a good correlation during speed step up change. Phase current as well as its reference command corresponding to this case are shown respectively in Figs. 8, 9. Figure 10 shows that three phase motor currents and their changes during speed step up.

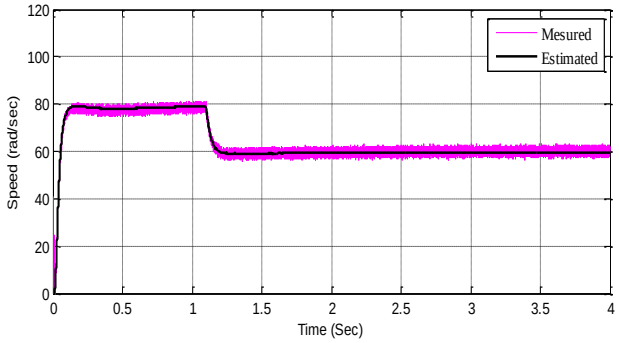


Fig 4.aEstimated and simulated motor speed signals obtained off line

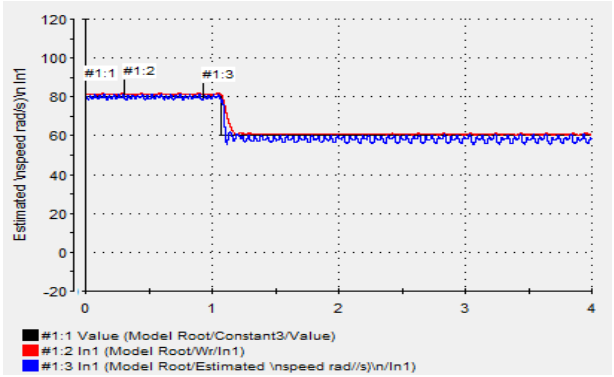


Fig 4.bEstimated and measured motor speed signals obtained in real time

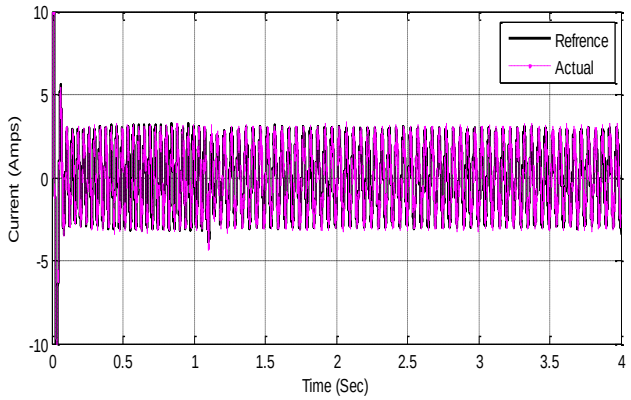


Fig. 5a (Simulation) Motor phase current and its reference command for step down of reference speed

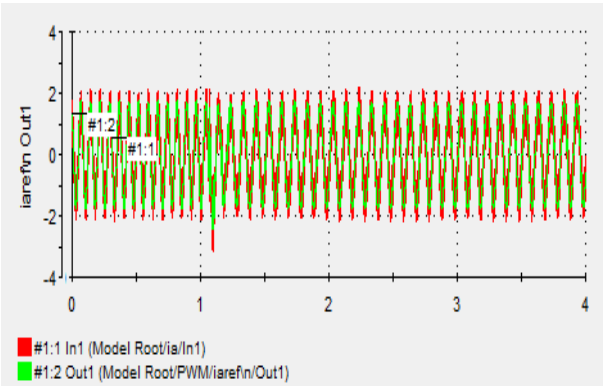
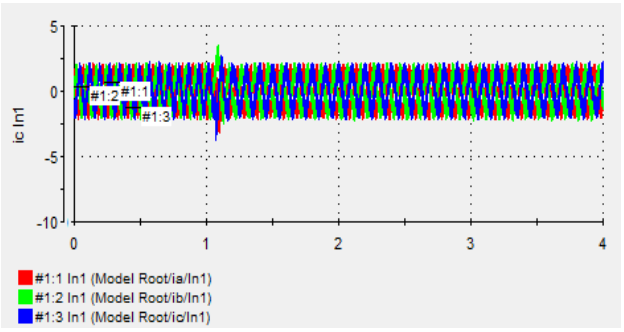


Fig. 6b (Experimental) Motor phase current and its reference command for step down of reference speed



]]

Fig. 6.a (Experimental) Motor phase current and its reference command for step down of reference speed

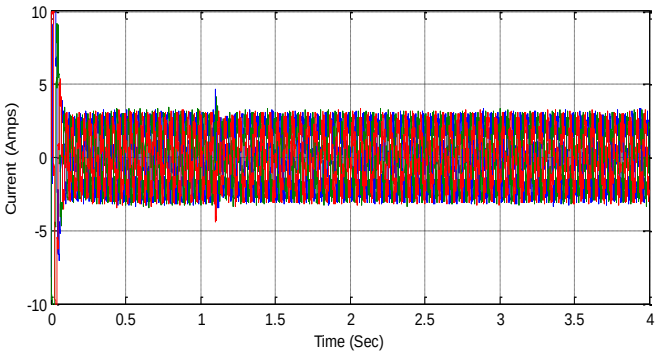


Fig. 6.b (Simulation) Motor three phase current for step down of reference speed

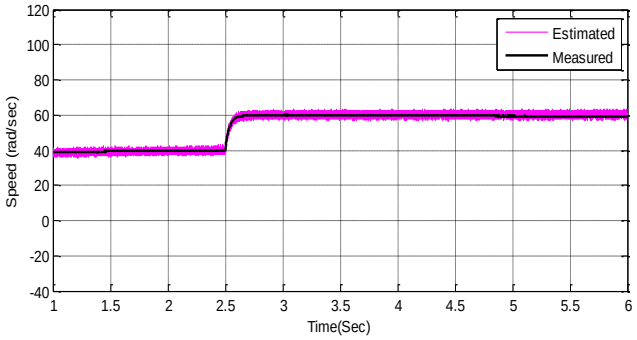


Fig 7.a Estimated and simulated motor speed signals obtained off line

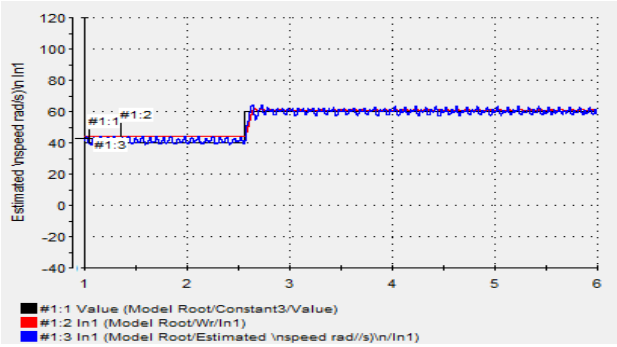


Fig 7.b Estimated and simulated motor speed signals obtained in real time

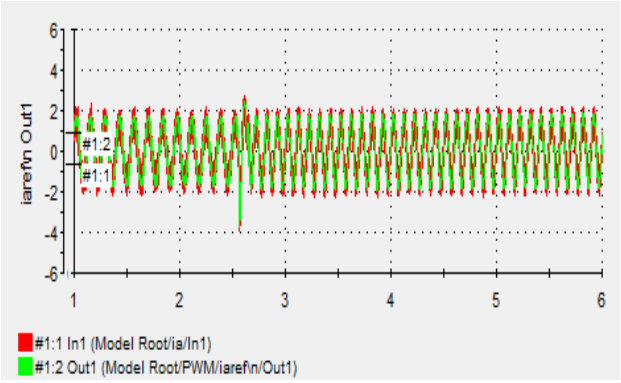


Fig 8.a (Experimental)Motor phase current and its reference command for step up of reference speed

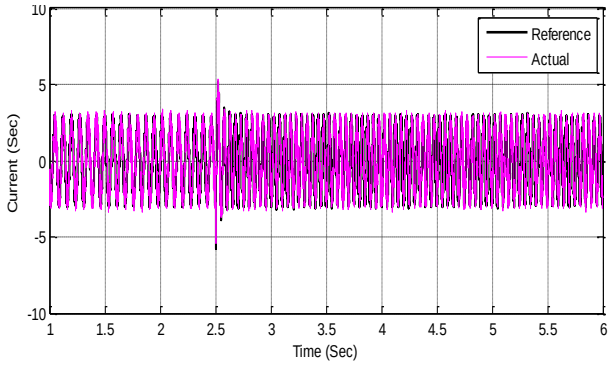


Fig. 8.b (Simulation) Motor phase current and its reference command for step up of reference speed

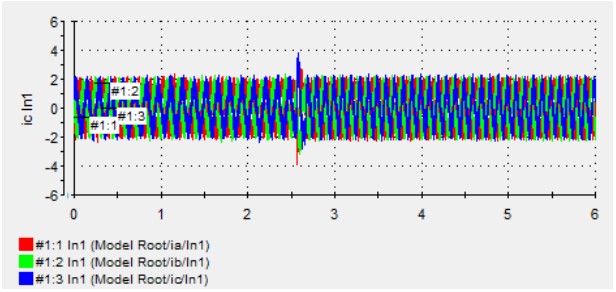


Fig. 9.a (Experimental) Motor three phase current step up of reference speed

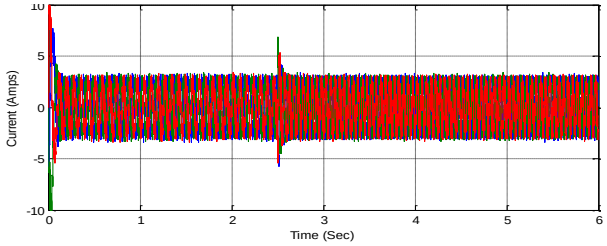


Fig. 9.b (Simulation) Motor three phase current for step up of reference speed

V. CONCLUSIONS

This paper presents the implementation of a speed sensorless induction motor drive system. PI Speed controller ofan indirect vector controlled induction motor drive system is based on cloud computing, whereas feedback speed signal is estimated using MRAS method. The drive system has been implemented based on MRAS speed estimation technique. The effectiveness of the proposed speed controller and speed estimation algorithm has been investigated under different operating conditions. A good correlation between simulated, estimated and measured speed signals has been obtained under different operating conditions. The results show the effectiveness and robustness of the proposed speed controller and speed estimation procedure.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[3] I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[8] J. Holtz, “Sensorerless Control of Induction Motor Drives,” Proceedings of IEEE, Vol. 90. No. 8, August 2002, PP. 1359-1394.

[9] [2] Faa-Jeng Lin, Rong-Jong Wai, and Pao-Chuan Lin, "Robust Speed Sensorless Induction Motor Drive," IEEE Transactions on Aerospace and Electronic Systems, Vol. 35, No. 2 April 1999, PP. 566-578.

[10] H. Kubota, I. Sato, Y. Tamura, K. Matsuse, Hisayoshi Ohta, and Y. Hori, "Regenerating Mode Low-Speed operation of Sensorless Induction Motor Drive With Adaptive Observer," IEEE-IA, Vol. 38, No. 4, July/August 2002, PP. 1081-1086.

[11] Young Ahn Kwon, and Sung Hwan Kim, "A New Scheme for Speed-Sensorless Control of Induction Motor," IEEE-IA, Vol. 51, No. 3, June 2004, PP. 545-550.

[12] M. Jemli, M. Boussak, M. Godda and M. B. A. Kamoun, "MRAS Identification Schemes for Sensorless Indirect Field Oriented Control of Induction Motor Drives with Rotor Resistance Tuning," Proc. of ICEM, Turkey 1998, PP. 1572-1577.

[13] M.N. Marwali, and A. Keyhani, "A comparative study of rotor flux based MRAS and back EMF based MRAS speed estimators for speed sensorless vector control of induction machines," IEEE-IAS Annual Meeting, 1997, PP. 160– 166.

[14] Z. M. Salem, M. M. Khater, S. A. Kalilah and S. A. Mahmoud, "Four Quadrant Speed Estimation Based on Model Reference Adaptive System," Engineering Research Journal, Vol.28, No.2, April 2005, PP. 169–177.

[15] T. C. Huang and M. A. El-Sharkawi, "High performance speed and position tracking of induction motors using multi-layer fuzzy control," IEEE-EC, Vol. 11, No 2, June 1996, PP. 353–358.

[16] B. Heber, L. Xu, and Y. Tang, "Fuzzy Logic Enhanced Speed Control of an Indirect Field Oriented Induction Motor Drive," IEEE-PE, Vol. 12, No. 5, Sept. 1997, PP. 772–778.

[17] Benoît Robyns, Frédérique Berthereau, Jean-Paul Hautier, and Hervé Buyse," A Fuzzy-Logic-Based Multimodel Field Orientation in an Indirect FOC of an Induction Motor," IEEE-IE, Vol. 47, No. 2, April 2000, PP. 380-388.

[18] Rong-Jong Wai, "Hybrid Control for Speed Sensorless Induction Motor Drive," IEEE Transactions on Fuzzy Systems, Vol. 9, No. 1, February 2001, PP. 116-138.

[19] Ximing Cheng and Minggao Ouyang, "Study of Speed Fuzzy Logic Real-Time Control System of Induction Traction Machine Based on a Single DSP Controller," Sixth International Conference on Electrical Machines and Systems, 2003 (IEEE Cat. No.03EX782), PP. 552-555.

[20] M.Abrate, G.Griva, F.Profumo, and A.Tenconi, "High Speed Sensorless Fuzzy-Like Luenberger Observer," Proceedings of the 30th Annual IEEE Power Electronics Specialists Conference, PESC'99, Charleston, USA, June 1999, PP. 477- 481.

[21] Chen C-Li and Chang M-Hui, "Optimal Design of Fuzzy Sliding Mode Control: A Comparative Study," International Journal on Fuzzy Sets Systems, Vol. 93, January 1998, PP. 37–48.

[22] Monaco, Ania (7 June 2012 [last update]). "A View Inside the Cloud". theinstitute.ieee.org (IEEE). Retrieved August 21, 2012.

[23] "Baburajan, Rajani, "The Rising Cloud Storage Market Opportunity Strengthens Vendors," infoTECH, August 24, 2011". It.tmcnet.com. 2011-08-24. Retrieved 2011-12-02.

[24] "Oestreich, Ken, "Converged Infrastructure," CTO Forum, November 15, 2010". Thectoforum.com. 2010-11-15. Retrieved 2011-12-02.

[25] "The NIST Definition of Cloud Computing". National Institute of Science and Technology. Retrieved 24 July 2011.

[26] Strachey, Christopher (June 1959). "Time Sharing in Large Fast Computers". Proceedings of the International Conference on Information processing, UNESCO. paper B.2.19: 336–341.

[27] Corbató, Fernando J.. "An Experimental Time-Sharing System". SJCC Proceedings. MIT. Retrieved 3 July 2012.

[28] "Cloud Computing: Clash of the clouds". The Economist. 2009-10-15. Retrieved

[29] "Gartner Says Cloud Computing Will Be As Influential As E-business". Gartner. Retrieved 2010-08-22.

[30] Gruman, Galen (2008-04-07)."What cloud computing really means". InfoWorld. Retrieved 2009-06-02.

[31] Sungkap Yeo and Hsien-Hsin S. Lee," Using Mathematical Modeling in Provisioning a Heterogeneous Cloud Computing Environment", IEEE Computer Society, AUGUST 2011 , pp55-62

APPENDIX

A. *Motor Data and Parameters:*

A three-phase, 4-pole, 380V, 50 Hz induction motor of the following parameters:

$R_s = 7.4826\ \Omega$ $R_r' = 3.6840\ \Omega$

$L_s = 0.4335$ H $L_r' = 0.4335$ H

$L_m = 0.4114$ H $J = 0.0200$ kg.m^2

B. *List of Principle Symbols;*

$$L_\sigma = L_s - \frac{L_m^2}{L_r},\quad T_r = \frac{L_r}{R_r},\quad k_1 = -\frac{R_s}{\sigma L_s} - \frac{L_m^2}{\sigma L_s L_r T_r} - \frac{1}{T_r}$$

$$k_2 = \frac{1}{\sigma L_s T_r},\quad k_3 = \frac{1}{\sigma L_s},\quad b_1 = \frac{1}{\sigma Ls},\quad \sigma = 1 - \frac{L_m^2}{L_s L_r}$$

V_{qse}, V_{dse}	qe-de –axis stator voltage
I_{qse} , I_{dse}	qe-de –axis stator current
λ_{qse} , λ_{dse}	qe-de –axis stator flux linkage
R_s , R_r	stator and rotor resistances
J, B	moment of inertia and viscous friction coefficients
L_s, L_r, L_m	stator, rotor and mutual inductances
T_e , T_L	electromagnetic and load torque

AUTHORS PROFILE

Dr. Z. M. Elbarbary: was born in Kaferelsheihk, Egypt, in 1971. He received the B.Sc.,M.Sc., and Ph.D. degrees in electrical engineering from Menoufiya University, Shebin El-Kom, Egypt, in 1994, 2002, and 2007, respectively. In 2009, he joinedKaferelsheihk University as an Assistant Professor. His fields of interests are ac motor drives; powerelectronics and Solar energy.

Dr. M. A. Abbas: Obtained his B.sc in electronics engineering from Mansoura university, faculty of engineering, Egypt,1998, M.sc degree in computer engineering from Mansoura University, 2002 and Ph.D degree in computer engineering from Mansoura University 2008. Since this date he worked as an assistant professor in Delta university, Faculty of engineering in department of communications and computer engineering. He is currently working in department of electrical engineering, King Khaled University, Abha , KSA. His area of interest is intelligent systems, medical informatics and bioinformatics.

Prediction of Compressive Strength of Self compacting Concrete with Flyash and Rice Husk Ash using Adaptive Neuro-fuzzy Inference System

S. S. Pathak, Dr. Sanjay Sharma, Dr. Hemant Sood
Research Scholar, Head of Deptt, Associate Professor
Civil Engineering Department
N.I.T.T.T.R, Chandigarh

Dr. R. K. Khitoliya
Professor and Head: Civil Engineering Department
Punjab Engineering College
Chandigarh, India

***Abstract*— Self-compacting concrete is an innovative concrete that does not require vibration for placing and compaction. It is able to flow under its own weight, completely filling formwork and achieving full compaction even in congested reinforcement without segregation and bleeding. In the present study self compacting concrete mixes were developed using blend of fly ash and rice husk ash. Fresh properties of theses mixes were tested by using standards recommended by EFNARC (European Federation for Specialist Construction Chemicals and Concrete system). Compressive strength at 28 days was obtained for these mixes. This paper presents development of Adaptive Neuro-fuzzy Inference System (ANFIS) model for predicting compressive strength of self compacting concrete using fly ash and rice husk ash. The input parameters used for model are cement, fly ash, rice husk ash and water content. Output parameter is compressive strength at 28 days. The results show that the implemented model is good at predicting compressive strength.**

Keywords- Self compacting concrete; ANFIS; Flyash.

I. INTRODUCTION

With growing population, industrialization, urbanization and globalization, there is corresponding growth in the demand for infrastructure. During the 20th century, concrete has emerged as the material of choice for modern infrastructural needs. It has occupied a unique position among modern construction materials. It gives considerable freedom to the architect to mould structural elements to any shape. Almost all concretes rely critically on being fully compacted. Insufficient compaction dramatically lowers ultimate performance of concrete inspite of good mix design. As concrete is produced and placed at construction sites, under conditions far from ideal, it often ends up with unpleasant results [1].

Concrete that is capable of compaction under its own weight and can occupy all the spaces in the forms, which self-levels, does not segregate and does not entrap air is termed as self-compacting concrete (SCC). For concrete to be self-compacting it should have filling ability, passing ability and resistance against segregation. Self compactability is obtained by limiting the coarse aggregate content and using lower water–powder ratio together with super plasticizers (SP). In the present study self compacting concrete is developed using blend of flyash and rice husk ash.

The artificial intelligence techniques have been used by many researchers to predict properties of concrete. M.C.Nataraja, M.A.Jayaram and C.N.Ravikumar developed a Fuzzy-Neuro model for normal concrete mix design. Model has been developed for approximate proportioning of standard concrete mixes [2]. B.K. Raghu Prasad, Hamid Eskandari and B.V. Venkatarama Reddy used artificial neural network to predict compressive strength of self compacting concrete and high performance concrete with high volume fly ash. The ANN was trained by data available in literature and validated by experimental results [3]. Mehdi Neshat, Ali Adeli, AzraMasoumi and Mehdi Sargolzae have carried out a comparative study on Adaptive Neuro-fuzzy Inference System (ANFIS) and Fuzzy Expert System Models (FIS) for concrete mix design [4]. Comparison between two systems FIS and ANFIS results showed that results of ANFIS system are better than FIS. AbdulkadirCüneytAydin, AhmetTortum and Muratyavuz developed a model for prediction of concrete elastic modulus using different models. Results of study indicated that the proposed ANFIS modeling approach outperforms the other given models in terms of prediction capability [5]. They proved that ANFIS approach is a viable tool for modeling the elastic modulus, as it results in more accurate predictions. ANFIS is one of such hybrid neuro-fuzzy inference expert systems which have been used in the present study.

Rafat Siddique, Pratibha Aggarwal and YogeshAggarwal predicted compressive strength of self-compacting concrete containing bottom ash using artificial neural network. The model developed from literature data was successfully extended to the experimental data [6]. Paresh Chandra Deka and Somanath N Diwate predicted 28-day compressive strength of Ready Mix Concrete by using soft computing techniques Artificial Neural Network (ANN) and Adaptive Neuro Fuzzy Inference System (ANFIS) modeling. ANFIS model having Gaussian membership function to predict concrete strength was found better than ANN [7].

II. ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM

Adaptive Neuro-fuzzy Inference System (ANFIS) is a multi-layer adaptive network-based fuzzy inference system proposed by Jang.

ANFIS as a modeling system consists of three distinct segments: i) the input parameters and membership functions, ii) the adaptive neuro-fuzzy inferencing system, iii) the output parameter and the defuzzifier. ANFIS architecture is represented in Fig.1. It consists of five layers of nodes. Out of the five layers, the first and the fourth layers consist of adaptive nodes while the second, third and fifth layers consist of fixed nodes. The circular nodes represent nodes that are fixed whereas the square nodes are nodes that have parameters to be learnt. Each of the input parameters has number of membership functions [8].

The process flows from layer 1 to layer 5. It is started by giving a number of sets of crisp values as input to be fuzzyfied in layer 1, passing through inference process in layer 2 and 3 where rules are applied, calculating output for each corresponding rules in layer 4 and then in layer 5 all outputs from layer 4 are summed up to get one final output. The main objective of the ANFIS is to determine the optimum values of the equivalent fuzzy inference system parameters by applying a learning algorithm using input-output data sets. The parameter optimization is done in such a way during training session that the error between the target and the actual output is minimized. Parameters are optimized by hybrid algorithm which combination of least square estimate and gradient descent method. The parameters to be optimized in ANFIS are the premise parameters which describe the shape of the membership functions, and the consequent parameters which describe the overall output of the system. The optimum parameters obtained are then used in testing session to calculate the prediction [9].

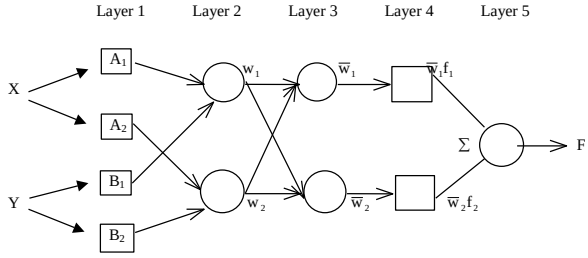


Figure 1. Anfis Structure of Two Rule Surgeno System

A Two Rule Sugeno ANFIS has rules of the form:

Rule 1: If x is A_1 and y is B_1, then $f_1 = p_1x + q_1y + r_1$

Rule 2: If x is A_2 and y is B_2, then $f_2 = p_2x + q_2y + r_2$.

where x and y are the inputs, Ai and Bi are the fuzzy sets, fi are the outputs within the fuzzy region specified by the fuzzy rule, p_i, q_i and r_i are the design parameters that are determined during the training process. In present study ANFIS model was developed using the fuzzy logic toolbox available in MATLAB software.

III. Development Of Anfis Model

Present experimental investigation was carried out to develop self compacting concrete (SCC) using blend flyash and Rice husk ash. It is important to mention that none of the test method for Self compacting concrete have yet been standardized and included in Indian Standard Code for the present. European guidelines (EFNARC) for testing, covers number of parameters ranging from material selection, mixture designs and testing methods like Slump flow test, L-box test, V-funnel test, U-box test, Orimet test and GTM screen stability for determining properties of SCC in fresh state. Most of Indian researchers are following these guidelines to determine the rheological properties of SCC mixes.

In present experimentation OPC 43 grade cement, flyash from Guru Gobind singh Super Thermal Power Station, Ropar, India and Rice Husk Ash (RHA) from Punjab Industrial area are being used for experimental investigation. A poly carboxylic based ether based superplasticizerGlenium B233 has been used. Total powder content i.e. cement fly ash and rice husk ash, was kept constant at 600 kg/m^3. The cement content was replaced by varying proportion of flyash and Rice husk ash. Total forty nine mixes were investigated for slump flow test, V-funnel test, U box test, L-Box, Orimet test and GTM Screen stability test. These tests were carried out as per EFNARC (European Federation for Specialist Construction Chemicals and Concrete system) standards. All these mix satisfied acceptance criteria laid down by EFNARC for fresh properties of self-compacting concrete. Compressive strength at the ages of 28 days was obtained. The ANFIS model is able to predict only one output parameter though input parameters may be many in number. In present study ANFIS model is developed to predict compressive strength at the ages of 28 days

A. Design of ANFIS Model

The ANFIS MODEL is designed by loading data, generating fuzzy inference system (FIS) and training FIS.

The input parameters were cement, flyash, rice husk ash and water content in kg/m^3. The Output parameter was standard 28-days cube strength in MPa. The data set having these four inputs and one target or output for 37 mixes presented in Table I was used for designing model and data set for remaining 12 mixes was used to check accuracy of prediction (Table III).

FIS was generated by loading data sets from Table I using Grid partition method. The performance of particular membership functions is good for certain data patterns. ANFIS models were generated, using different functions like triangular (trimf), trapezoidal (trapmf), generalized bell-shaped (gbellmf) andgaussian curve (gaussmf) membership function by conducting trial runs. From this a membership function was selected. The number of membership functions was three per parameter.

TABLE I. Training Data

Mix No	Cement (kg/m^3)	Flyash (kg/m^3)	RHA (kg/m^3)	Water (kg/m^3)	Compressive Strength (MPa)
1	240	360	0	228.9	28.52
2	240	342	18	229.5	27.95
3	240	306	54	230.9	26.81
4	240	270	90	234	24.24
5	240	262	108	248	22.3
6	300	300	0	218.5	32.24
7	300	285	15	219.3	31.6
8	300	255	45	221	30.31
9	300	240	60	223.5	29.34

Mix No	Cement (kg/m^3)	Flyash (kg/m^3)	RHA (kg/m^3)	Water (kg/m^3)	Compressive Strength (MPa)
10	300	225	75	224.3	28.37
11	300	210	90	225.1	27.4
12	360	240	0	208.8	39.68
13	360	228	12	209.8	38.89
14	360	204	36	211.7	37.3
15	360	180	60	216.1	34.92
16	360	168	72	217.6	33.73
17	420	180	0	200	46.38
18	420	171	9	201	45.45
19	420	153	27	203.1	43.6
20	420	135	45	208	40.81
21	420	126	54	209.7	39.42
22	450	150	0	195.8	48.98
23	450	127.5	22.5	199.1	46.04
24	450	112.5	37.5	204.2	43.1
25	450	105	45	205.9	41.63
26	480	120	0	191.9	51.58
27	480	114	6	192.9	50.55
28	480	102	18	195.2	48.49
29	480	96	24	198.7	46.94
30	480	90	30	200.5	45.39
31	480	84	36	202.4	43.84
32	510	90	0	188.1	54.19
33	510	85.5	4.5	189.2	53.11
34	510	76.5	13.5	191.5	50.94
35	510	67.5	22.5	196.9	47.69
36	510	63	27	198.8	46.06
37	600	0	0	178	65

Error obtained for various member functions is presentedin Table II

TABLE II. ERROR FOR MEMBERSHIP FUNCTIONS

Membership Function	*trimf*	*trapmf*	*gbellmf*	*gaussmf*
Error	0.017	0.881	0.0321	0.0213

The present data shows minimum error levels for triangular input membership function.

Hence triangular membership function along with three parameters was used for the present study. The triangular membership functions for cement content is shown in Fig. 2

The output membership function can either be a constant membership function or a linear membership function. For the present data, constant output membership function produced minimum error. Fuzzy inference system is trained by hybrid network for 50 numbers of epochs.

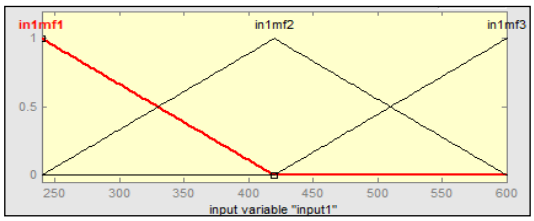


Figure 2. Membership fuction of Cement

The optimization methods train membership function parameters to emulate the training data. The hybrid optimization method is a combination of least-squares and back propagation gradient descent method. In hybrid method, model tunes with forward pass and backward pass [4]. Training error tolerance was set to zero.

The model having 81 fuzzy rules is created and Details of various parameters obtained after training are as given below

- Number of nodes : 193
- Number of linear parameters : 81
- Number of nonlinear parameters : 36
- Total number of parameters : 117
- Number of training data pairs : 37
- Number of fuzzy rules : 81
- Error : 0.01709

Structure of ANFIS model has been shown in Fig. 3.

B. Testing of ANFIS Model

Data set of four inputs for 12 mixes was used to predict compressive strength of mix in order to check accuracy of prediction of a model. Predicted compressive strength by a model and percentage error as compared to actual compressive strength is presented in Table III.

The surface viewer of compressive strength (output) with cement (input 1) and flash content (input 2) is shown in Fig. 4.

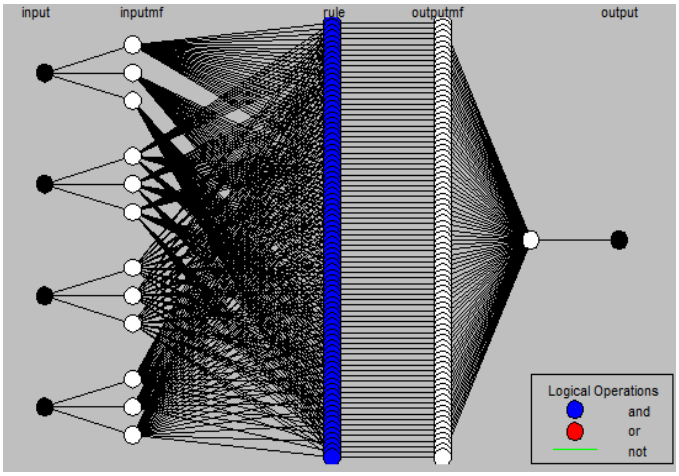


Figure 3. Structure of ANFIS Model

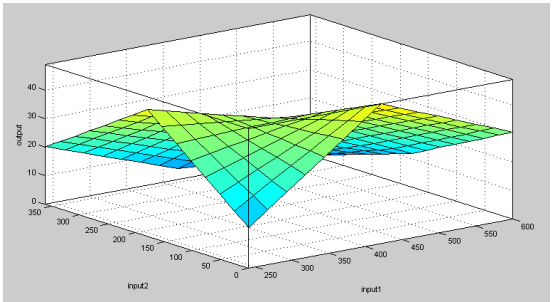


Figure 4. Surface of Compressive Strength for Cement and Flyash Content

TABLE III. ERROR ANALYSIS IN TESTING OF ANFIS MODEL

Sr. No	Input Data				Compressive Strength (In MPa)		% Error
	Cement	*Flyash*	*RHA*	*Water*	*Predicted*	*Actual*	
1	240	324	36	230.3	27.4	27.38	-1.90
2	240	288	72	233	25.8	25.1	5.18
3	300	270	30	220.1	31	30.95	-0.48
4	360	216	24	210.7	38.1	38.09	-0.03
5	360	192	48	214.7	35.5	36.11	-0.25
6	420	162	18	202	44.5	44.52	0.27
7	420	144	36	206.3	42.2	42.21	0.73
8	450	135	15	198	47.1	47.02	0.89
9	450	120	30	202.5	44.7	44.57	-1.86
10	480	108	12	194.1	49.6	49.52	0.65
11	510	81	9	190.4	52.1	52.02	0.04
12	510	72	18	195.1	49.2	49.31	0.22
						Average Error	0.29%

IV. RESULTS AND DISCUSSIONS

The results of compressive strength predicted by the model when compared with experimental results found average error of 0.29 % only (Table III).

Predicted compressive strength plotted against actual compressive strength in Fig 5 show very good coefficient of correlation 0.91.ANFIS model shows the excellent performance and is capable to predict compressive strength.

The model for prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural network was developed by Rafat Siddique, Pratibha Aggarwal and Yogesh Aggarwal for the data from literature. Correlation coefficient of 0.91 was achieved for prediction of compressive strength at 28 days.

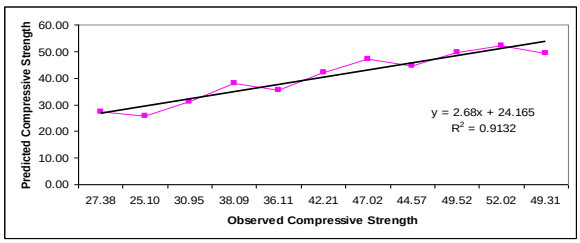


Figure 5. Correlation between Actual and Predicted Compressive Strength

The overall excellent performance of model depicted in Fig. 5.

V. CONCLUSION

The following conclusions can be drawn from this work:

- The ANFIS having the Triangular membership function could predict the 28-day compression strength of self compacting concrete with satisfactory performance.
- From the results obtained it can be concluded that the ANFIS models are more suitable in modeling of complex problems and save a lot of computational effort. The use of these networks will help in solving more complex problems.

REFERENCES

[1] M. Shobha, D. Harish Mohan, P. S. N. Raju,"Aggregate size and behaviour of self-compacting concrete",Proceedings of the ICE - Construction Materials",Volume 159, Issue 4, 01 November 2006, pp. 147 –152

[2] G M.C.Nataraja, M.A.Jayaram, C.N.Ravikumar, "A Fuzzy-Neuro Model for Normal Concrete Mix Design", Engineering Letters,13:2, EL_13_2_8, 4 August 2006.

[3] B.K. Raghu Prasad, Hamid Eskandari,B.V.Venkatarama Reddy, "Prediction of compressive strength of SCC and HPC with highvolume fly ash using ANN", Construction and Building Materials 23 (2009), pp.117–128.

[4] Mehdi Neshat, Ali Adeli , Azra masoumi , Mehdi sargolzae, "A Comparative Study on ANFIS and Fuzzy Expert SystemModels for Concrete Mix Design",International Journal of Computer Science Issues, Online Vol. 8, Issue 3, No. 2, May 2011

[5] Abdulkadir Cüneyt Aydin, Ahmet Tortum and Muratyavuz, "Prediction of concrete elastic modulus using adaptiveneuro-fuzzy inference system", Civil Engineering and Environmental Systems ,Vol. 23, No. 4, December 2006, pp. 295–309

[6] Rafat Siddique ,Paratibha Aggarwal , Yogesh Aggarwal, "Prediction of compressive strength of self-compacting concrete containing bottom ash using artificial neural networks",Advances in Engineering Software (2011), pp 780–786

[7] Paresh Chandra Deka and Somanath N Diwate, "Modeling Compressive Strength of Ready Mix Concrete Using Soft Computing Techniques", International Journal of Earth Sciences and Engineering , October 2011, pp. 793-796

[8] Balasubramaniam, V., P.N. Raghunath and K. Suguna, "An Adaptive Neuro-Fuzzy Inference System Based Modeling for Corrosion-Damaged Reinforced HSC Beams Strengthened with External Glass Fibre Reinforced Polymer Laminates", Journal of Computer Science 8 (6): ,pp. 879-890,

[9] Vidi Bhuwana, "Rainfall Runoff Modeling by Using Adaptive-Network-BasedFuzzy Inference System (ANFIS) - Case Study Ciliwung River (On line Paper)"

AUTHORS PROFILE

Mr. S. S. Pathak is pursuing Ph. D program at Civil Engineering Department, National Institute of Technical Teachers Training and Research, Chandigarh under Punjab University. He has teaching experience more than 20 years. His areas of research include self compacting concrete using different mineral admixtures.

Dr. Sanjay Kumar Sharma is professor and head of Civil Engineering Department, National Institute of Technical Teachers Training and Research, Chandigarh. He has more than 25 years of experience in teaching. His areas of research includeEnvironmental Engineering, Building Repair and Rehabilitation, Irrigation and Hydraulics, Public Health Engg. and Concrete Technology.

Dr. HemantSood is associate professor in Civil Engineering Department, NIITTTR, Chandigarh. He has also served in Rail India Technical and Economic Services Limited (RITES), under Ministry of Railways as Assistant Manager. He has more than 22 years of teaching and industrial experience. His areas of research include Highway Engineering, Pavement Design, Concrete Mix Design and Concrete Technology.

Dr. R. K. Khitoliyaobtained M. Tech (Civil Engg.) and Ph.D (Civil Engg.) from Indian Institute of Technology, Delhi. He is at present working as Professor and Head of Civil Engineering, Punjab Engineering College, Chandigarh. He is ex-director of Harcourt Butlar Technological Institute, Kanpur. His areas of interest are Geotechnicalinvestigations,Environmental Legislation, Environmental Management, Sustainable Development, Environmental Impact Assessment and Concrete Technology.

Smart Card Based Integrated Electronic Health Record System For Clinical Practice

N. Anju Latha*, B. Rama Murthy
Department of Electronics
Sri Krishnadevaraya University
Anantapur, A.P, INDIA

U. Sunitha
Department of Electronics
Sri Krishnadevaraya University
Anantapur, A.P, INDIA

***Abstract*— Smart cards are used in information technologies as portable integrated devices with data storage and data processing capabilities. As in other fields, smart card use in health systems became popular due to their increased capacity and performance. Smart cards are used as a Electronic Health Record (EHR) Their efficient use with easy and fast data access facilities leads to implementation particularly widespread in hospitals. In this paper, a smart card based Integrated Electronic health Record System is developed. The system uses smart card for personal identification and transfer of health data and provides data communication. In addition to personal information, general health information about the patient is also loaded to patient smart card. Health care providers use smart cards to access data on patient cards. Electronic health records have number of advantages over the paper record, which improve the accuracy, quality of patient care, reduce the cost, efficiency, productivity. In present work we measure the biomedical parameters like Blood Pressure, Diabetes Mellitus and Pulse oxygen measurement.,etc clinical parameters of patient and store health details in Electronic Health record. The system has been successfully tested and implemented (Abstract)**

Keywords- Electronic health record; Smart card technology; Healthcare using smart cards.

I. INTRODUCTION

Automation systems in hospitals serve the purpose of providing an efficient working environment for health care professionals. Access to accurate health data quickly is one of the main functions of this system. There can be many sources that the information related to the patients can be obtained from the patient, test results, doctor diagnoses for patient illness, health measurement devices and previously stored patient information [1]. The usual way of obtaining relevant data is from paper record. The Paper-based records have a low cost and have limitations such as difficult to access, time-consuming to update, secure, impossible to share and maintain for lifelong.

The problems can be solved by increasing the capabilities of hospital automation systems by using intelligent storage and retrieval mechanism. Smart card can play a key role in sharing patient specific information. The patient can carry the health smart card with him/her anywhere and anytime and present it to the doctor at the time of consultation. Smart cards are more suitable to use in health care information systems because of they are cheap, easy to use, carry and update with new information and should not get damaged easily.

Smart cards can be described as portable integrated devices that store and process data. These tiny computers with their own memories and processors have a widespread usage especially in telecommunication and mass transit systems [2]. Speed, security and portability properties make smart cards a potential tool in healthcare systems. Many countries implement or continue to develop such systems including smart card components. An intensive study on these systems is seen in working projects such as Sesam Vitale in France and DENT card in Germany [3]. Also European Commission supported France–Belgium project Transcards [4], EU IV. Framework R&D project Netlink [5]. The Health Insurance Card project of Slovenia provides nation-wide use of smart cards in health sector [6, 7].

The objective of the present work is to develop a multifunctional user-friendly biomedical measurement device of an Integrated Electronic Health Record System, which provides a complete e-record in place of paper record. The following health parameters are required by any doctor such as Blood Pressure, Blood glucose level measurement, pulse oxygen meter, clinical analyzer. Each device was designed, built, tested and calibrated separately. The modules devices are connected serially to a Personal Computer with a Visual Basic software package. The data acquired from each system is then displayed on the PC monitor and store data in Health Smart card [8].

II. GENERAL SPECIFICATIONS

The system developed is called Integrated Electronic Health Record System (IEHRS).Patients have Health smart cards in IEHRS. Doctors use these cards to access health data with database. Smart cards are used as a mobile health data carrier.

In patient card, personal and patient information is stored. In personal information patient ID, patient's name, surname, birth date, blood type, gender, address and mobile telephone numbers are stored in patient card. Patient health information stored in the card is chronic and/or important former diseases with diagnosis dates, permanently used medications with doses, allergies with diagnosis dates, immunizations with their dates, surgical operations including operation date, clinic name and summary information. Patient's last examination and prescription information are also stored in card [9].

Each Computer has a connected card acceptance devices (CAD) and they can connect to databases. When a doctor

inserted a card in CAD patient's health data are displayed [10]. After examination, doctor updates inspection and prescription information on patient smart card with new data. Hospital administration unit to record new inspection and prescription data stored on smart card. The hospital administration manage hospital database as well as the other responsibilities of administration is to record new patients to system and to perform smart card related operations like smart card preparation for new users and data update from smart cards to database and database to smart cards[11].

III. System Description- IEHRS Block Diagram & Description

The Integrated Electronic Health Record System (IEHRS) block diagram as shown in figure 1. The primary function of the device is to operate as a medical pre-screening/diagnostic device. The first stage lists the four different types of medical data to be measured by the device. Blood Pressure, Blood Glucose, Pulse Oxymeter and clinical analyzer. In the second stage of the system a Personal Computer reads data from these medical devices and stores it in a specified patient database. In the third stage the stored data is transferred to a health smart card using smart card reader/ writer to create an e-record.the graphical user interface package has developed using Visual Basic for the present work integrated electronic health record system.

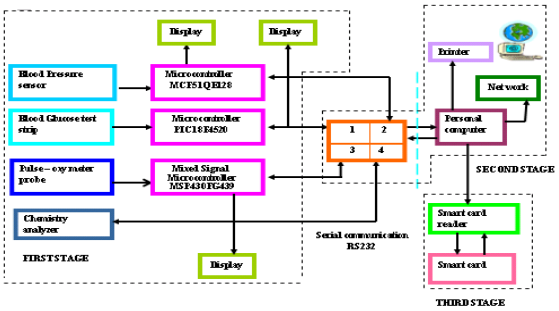


Figure 1: Block Diagram of Integrated Electronic Health Record System

IV. Hardware

A. Blood Pressure Meter

Blood pressure is the force exerted by circulating blood on the walls of blood vessels. BP is the force created by the heart as it pushes blood into the arteries through the circulatory system. A typical reading will look like "120/80" and is measured in mmHg. The upper value is the Systolic Blood Pressure, and the lower value is the Diastolic Blood Pressure.

In the present study, a microcontroller MCF51QE128based system for the measurement of blood pressure is developed. Further, an LCD module to display the data of systolic and diastolic blood pressure and MR2A16A asynchronous magneto resistive RAM to store measured values of systolic diastolic pressure values are interfaced with the microcontroller, which reduces the hardware complexity.

Software is developed in C language using IAR embedded work bench IDE due to the inherent language flexibility, the extent of support and its potential for portability across a wide range of hardware. The blood pressure measurements were tested and the results are shown in Table 1.

TABLE 1: Measured Blood Pressure values (mmHg)

S. No	Sphygmomanometer	Omron HEM-735C	Present meter
1	120/80	118/78	119/79
2	122/80	120/82	122/80
3	130/90	128/88	129/90
4	150/90	149/89	149/92
5	160/100	158/100	158/102

A. Blood Sugar Level Meter

Glucose is one of the body's main sources of energy. Diabetes is a chronic metabolic disorder characterized by a high concentration of sugar in the blood. Diabetes mellitus is a condition in which the pancreas no longer produces enough insulin or when cells stop responding to the insulin that is produced, so that glucose in the blood cannot be absorbed into the cells of the body. The body maintains blood glucose levels within a narrow range (70-130mg/dl).

In the present design, a microcontroller based system for the measurement of blood glucose is designed and developed. It is based on the Amperometric method. A PIC 18F4520 microcontroller is used in the present study. LCD module is used to display measured values of blood glucose. The MAX232 is a dual line driver/receiver, converts signals form an RS-232 serial port to TTL compatible signals is interfaced with the microcontroller. Software is developed in C language using MPLAB IDE for the Microchip Technology. The blood glucose measurements were tested and the results are shown in Table 2.

TABLE 2: Measured Blood Glucose values (mg/dl)

S. No	One touch Ultra 2	Clinical Analyzer RT1904C	Present meter
1	136	134	131
2	149	148	152
3	183	182	180
4	315	317	321
5	337	332	339

B. Oxygen Saturation (SaO2)

Oxygen saturation level or, SaO2 reveals the amount of oxygen carried by the blood. A low SaO2 level can indicate that the body is not getting enough oxygen, and could be a symptom of lung or heart disease. An overly high SaO2 level can cause oxygen poisoning. Pulse Oxymeter is based on the fractional change in light transmission during an arterial pulse at two different wavelengths.

TABLE 3: Pulse oxygen measurement

S. No	BPL Meter		Present Meter	
	Pulse oxygen	*Pulse rate*	*Pulse Oxygen*	*Pulse rate*
1	98	68	97	67
2	99	68	99	68
3	99	70	99	71
4	98	67	97	67
5	98	69	98	68

In present design of Pulsoximeter MSP430FG437 microcontrolleris used. The pulsoximeter consists of a peripheral probe combined with the MCU displaying the oxygen saturation and pulse rate on LCD glass. The probe is placed on a peripheral point of the body such as a fingertip, ear lobe or the nose. The probe includes two light emitting diodes (LEDs), one in the visible red spectrum (660 nm) and the other in the infrared spectrum (940 nm). The percentage of oxygen in the body is worked by measuring the intensity from each frequency of light after it transmits through the body and then calculating the ratio between these two intensities. The pulse oxygen meter measurements values are tested and the results are shown in Table 3.

The above instruments are tested and results are found to be satisfactory. The instruments are handheld, rugged, low cost, low energy consumption, wearable and cost effective compared to the other commercially available**.** The measured values are transmitted to a Personal Computer and stored in the patient database to create an e- record

B. Clinical Chemistry Analyzer - RT1904C

Clinical chemistry analysis is one of the most important areas within clinical laboratory analysis. The term clinical chemistry usually refers to determining the concentration or activity of a protein, carbohydrate, lipid, electrolyte, enzyme or small molecule in easily-collected bodily fluids such as blood, serum, plasma or urine. However, it is not necessarily limited to these determinations. In the present design an RT1904C Clinical Chemistry Analyzer is used and the measurement values of all clinical parameters are transferred to EHR system interfacing with PC through RS-232 with developed IEHRS software package.

V. SOFTWARE

In IEHRS we developed exclusive software in Visual Basic for the reading and writing of the health information on to electronic health recording system.

The GUI program in Visual Basic is used to communicate with the medical devices. The code allows us to send and receive characters over RS232 with SQL support for a user database in which the patient data can be stored securely for usage. The patient database could keep track of medical checkup results of patients and are useful in establishing health histories and monitoring trends. Periodically, the database can be uploaded to any specialist for further diagnosis, prevention, and treatment. For example, mild diabetics only require blood sugar level monitoring a few times a week. A database can keep track of their blood sugar history, revealing trends and help develop a plan for controlling their disorder.

When a file is opened it displays the menu bar which consists of six items having individual functions Reader, EHR_Administration, Doctor, Lab and Medical. The *Reader* option on the menu consists of Connect, Disconnect. The CAD can be connected or disconnected with the computer by selecting the *Connect or Disconnect.* The option *EHR_Administration* consists of Registration, admission, discharge, and transfer details of a patient provides vital information for accurate patient identification and assessment, including chief complaint, patient disposition, etc. The personal details contain the personal details, emergency details and insurance details of a patient. The emergency details contain the blood Group, allergies of a patient... etc., The insurance details of a patient consist of a patient insurance policy name, number, policy type, date of issue, date of expiry and amount of patient. The Doctor option consists of Consultation, Discharge Process and Prescription. Laboratory System Components consists of integrate orders, results from laboratory instruments, schedules, billing and other administrative information. In the present work measurement devices like Blood Pressure meter, Blood Glucose meter, Pulseoxy meter and clinical analyzer have been integrated. The measured devices are communicated with the Personal Computer by selecting the serial port. The results are transmitted from the electronic medical device to the Personal Computer. And further it will be transferred to health smart card. Pharmacy System Components consists of the patient's name, number and cost details. The electronic prescribing consists of the drug details of a patient. Flowchart of the system is as shown in fig 2.

VI. FLOW CHART

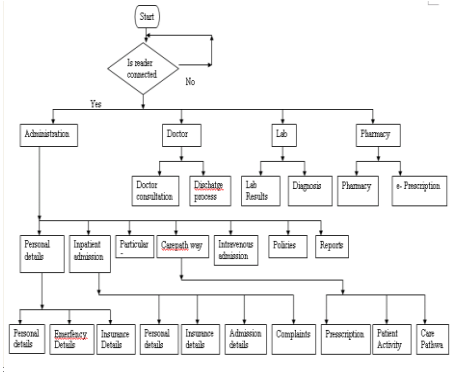


Figure 2: Flow chart of IEHR System

VII. IMPLEMENTATION

Smart card software is developed in Visual Basic. ACR38 serial card read/write unit is used as CAD in system. It is connected to a terminal PC with 9600 baud data transmission rate. Card terminal interface packages are deployed to a PC with Intel PIII 650MHz CPU running MSWindows 2000 operating system. That PC represents a card terminal located in a doctor's room. System administration software is also deployed and tested on this computer.MS SQL Server 2000 is used both for hospital database implementations.

Some performance measurements are also obtained during system tests, containing elapsed time measurement during data transmission between smart card client applications and smart cards. Considering data bus with 9600 baud, to send a command and receive a response with 255 bytes of data and display content in related interface take approximately 1.5 s. To write 255 bytes of data to smart card and receive response from card takes approximately 2 s.

Screenshot in Fig. 3 is taken just after the doctor has accepted a patient and opened a patient session on the clinic application.

All displayed data (except the remote database message) is received from the patient's smart card. In Fig. 4 patient's medical information about his allergies, diseases, etc., are displayed. The doctor can also access other patient information (e.g. surgical operations, last inspection and prescription) stored on patient's smart card by using the other submenus of the "Patient" menu. The doctor has received her patient's health records according to the distributed object protocol proposed in this paper.

The doctor updates the patient's health records it is enough for her to simply press the "Update" button and approve the operation.

Figure 3: Patient details

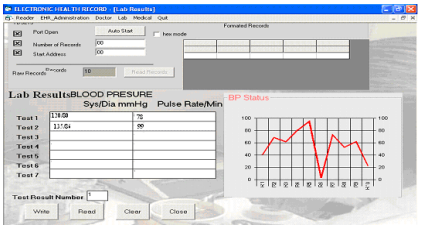


Figure 4: Test results

VIII. Status Report

IEHRS can be considered as a powerful healthcare automation with integration of smart card use into existing hospital information systems. Its distributed protocol enables mobile and secure access to the patient records and facilitates roles of both healthcare professionals and patients.

However, contribution of smart cards in those studies is limited even in the systems that are currently in use. System has a restricted design in which smart cards only behave as a portable health report card. Potential security and authorization features are not fully presented.

On the other hand, protocol introduced with IEHRS allows use of cards in data storage and but security has to introduced in addition to mobile data carriage.

It should also be noted that architecture of IEHRS takes care of the easy integration of the currently working health information systems in a hospital. Hence, users of the old information system (both health professionals and patients) can adapt to the new system easily and quickly. Finally, it presents electronic prescription data structure with working pharmacy software. Fig 5 shows the electronic prescription format.

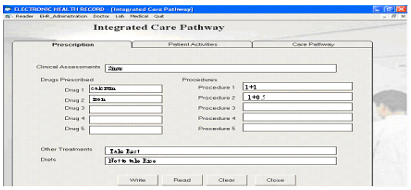


Fig 5: Electronic Prescription window

IX. Future Plans

A healthcare automation system based on smart cards is designed and developed. The most important problem encountered during system development is the lack of medical data store and retrieval standardization in healthcare sector. Existence of a worldwide-generic coding standard for healthcare data will surely ease the design and development of smart card based healthcare systems. Our system has its own specific medical data coding in databases. However as standardization occurs, the database can be redesigned to meet those standards. The capacity increase and cheaper costs will improve quality of smart card services.

The present patient cards of the IEHRS can only store last inspection and prescription data. One of our future plans is to integrate smart cards with higher capacity into the system to provide storage of more than one inspection and prescription data on card and simplify the doctor's examination process. With use of such high capacity smart cards, we also intend to store extra medical information like X-ray films and test documents on smart cards.

References

[1]. "Design and implementation of a smart card based healthcare information system", Geylani Kardas, E. Turhan Tunali, journal of Elsevier, computer methods and programs in biomedicine 8 1 (2 0 0 6) 66–78, 2005

[2]. Z. Chen, Java Card TM Technology for Smart Cards Architecture and Programmer's Guide, Addison-Wesley, MA, USA, 2000.

[3]. C. Pagetti, C. Mazini, M. Pierantoni, G. Gualandi, H.Schepel, A European Health Card Final Report, European Parliament, Directorate General for Research, Document for STOA Panel, 2001, pp. 16–29.

[4]. Transcards, GIE Sesam Vitale, Transcards Project, URL: http: //www. esamvitale.fr /html/projets/trans cards / tcd accueil eng.htm, last accessed: 2002.

[5]. Netlink, GIE Sesam Vitale, Netlink Project, URL:http:// www.sesamvitale.fr/ html /projets/netlink/index.htm, last accessed: 2002

[6]. R. Novak, G. Kandus, D. Trcek, Further development of a smart-card based health care information system in Slovenia, in: Presented at the Fifth International Congress on Conference and Exhibition on Cards Applications in Health Care: Health Cards'99, Milan, Italy, 1999.

[7]. R. Novak, G. Kandus, D. Trcek, Slovene smart-card and IP based health-care information system infrastructure, in: International Journal of Medical Informatics, vol. 61, Elsevier, 2001, pp. 33–43.

[8]. "Smart cards applications in the Healthcare Systems", claudiu oltean, Journal of Mobile, Embedded and Distributed Systems.

[9]. "Health Care Implementation by Means of Smart Cards", ,Dr. Magdy E. Elhen nawy, Dr. M. Amer, A. Abdelhafeez , International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011 ISSN (Online): 1694-0814 www.IJCSI.org

[10]. "Health Smart Home", Ahmad CHOUKEIR, Batoul FNEISH , Nour ZAAROUR , Walid FAHS , Mohammad AYACHE IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 6, November 2010, ISSN : 1694- 0814.

[11]. Alvin T. S. Chan, "smart card: towards a mobile health care management system", International Journal of Medical Informatics, Elsevier Science Ireland Ltd, Volume 57, Issues 2-3, Pages 127-137, July 2000.

AUTHORS PROFILE

Dr. N. Anju latha presently working as teaching faculty member in thd department of Instrumentation and USIC, Sri Krishnadevaraya University, Anantapur. She is having three years Industrial experience as R&D Engineer in conceptualization and development of Micro controller/Microprocessor-based products and solutions for Bio-Medical Instruments, Consumer Electronics and Smart Cards based devices.

Dr.B.RamaMurthy is presently working as a Professor in the Department of Instrumentation & USIC, Sri Krishnadevaraya University, Anantapur, Andhra Pradesh, India. He is having 20 years of Research & teaching experience. Under His guidance 8 Ph.D & 4 M.Phils are awarded. His areas of interest are Embedded Systems, Network and Mobile Communications, network security systems, Industrial and Bio-medical Instrumentation.

U.Suneetha is doing P.hd in the Department of Electronics and communications, Sri Krishna Devaraya university, Anantapur. She also working as a Teaching assistant in the Department of Electronics and communications, Sri Krishna Devaraya university, anantapur, Andhrapradesh, India. She having nine years of teaching experience. Her area of interest are Wireless communications and embedded systems

NF-SAVO: Neuro-Fuzzy system for Arabic Video OCR

Mohamed Ben Halima, Hichem karray,
Adel. M. Alimi
REGIM: REsearch Group on Intelligent Machines
University of Sfax, National School of Engineers (ENIS)
B.P 1173. 3038 Sfax-Tunisia

Ana Fernández Vila
Grupo SSI, Dept. Ingeniería Telemática
University of Vigo
Spain

***Abstract*— In this paper we propose a robust approach for text extraction and recognition from video clips which is called Neuro-Fuzzy system for Arabic Video OCR. In Arabic video text recognition, a number of noise components provide the text relatively more complicated to separate from the background. Further, the characters can be moving or presented in a diversity of colors, sizes and fonts that are not uniform. Added to this, is the fact that the background is usually moving making text extraction a more intricate process.**

Video include two kinds of text, scene text and artificial text. Scene text is usually text that becomes part of the scene itself as it is recorded at the time of filming the scene. But artificial text is produced separately and away from the scene and is laid over it at a later stage or during the post processing time. The emergence of artificial text is consequently vigilantly directed. This type of text carries with it important information that helps in video referencing, indexing and retrieval.

Keywords- Arabic Video OCR; Text Localization; Text Detection; Text extraction; Pattern Recognition; Neuro-Fuzy.

I. INTRODUCTION

Optical Character Recognition for Arabic scripts is almost a solved problem for document images and researchers are now focusing on extraction and recognition of Arabic text from video scenes. This new and promising field in character recognition is called Arabic Video OCR and has several applications similar to video annotation, indexing, retrieval, search, digital libraries, and lecture video indexing. Images extracted from video sequences are of low resolution. For this, the commercial OCR provides poor results. For example, both Sakhr and OmniPage performed poorly on this image. Sakhr achieved 38.88% accuracy and OmniPage achieved 35.79% accuracy.

Works on text extraction may be generally grouped into four categories [11]: First category is the connected component methods which detect text by extracting the connected components of monotone colors that obey certain size, shape, and spatial alignment constraints. The second is the texture methods which treat the text region as a special type of texture and employ conventional texture classification method to extract text. The third is the edge detection methods which have been increasingly used for caption extraction due to the rich edge concentration in characters *[8]*. Finally, the correlation based methods which use any kind of correlation in order to decide if a pixel belongs to a character or not. Recognition of text in video is more difficult than many other OCR applications (e.g., reading printed matter) because of degradation, such as background noise, and deformation, like the variation in fonts.

Our system is divided into two steps: first we extract the textual information from the video sequence and second we recognize this text. Text recognition, even when applied to lines potentially containing text, remains a difficult problem given the variety of fonts and colors and the presence of complex background. Recognition is addressed by a segmentation step followed by a step of optical character recognition in a framework of multiple hypotheses.

In Figure 1, we present some examples of text in video frames.

Figure 1. Examples of text in video frames

News programs are particular audio-visual documents they are generally formed by a set of semantically independent stories. For this reason before starting the extraction of the textual information from the news programs, segmentation into stories will be done. The following figure shows an example of image taken from video broadcast of Aljazeera TV, Tunisia 1 TV, France 24 TV and Alarabiya TV.

The rest of the paper will be organized in four sections. In the section 2, we discuss works related to news segmentation. In section 3, we will present how we detect and localize the textual information in video sequence. In section 4, we will present how we extract the text. In

section 5, we will present how we recognize it. We conclude with directions for future work.

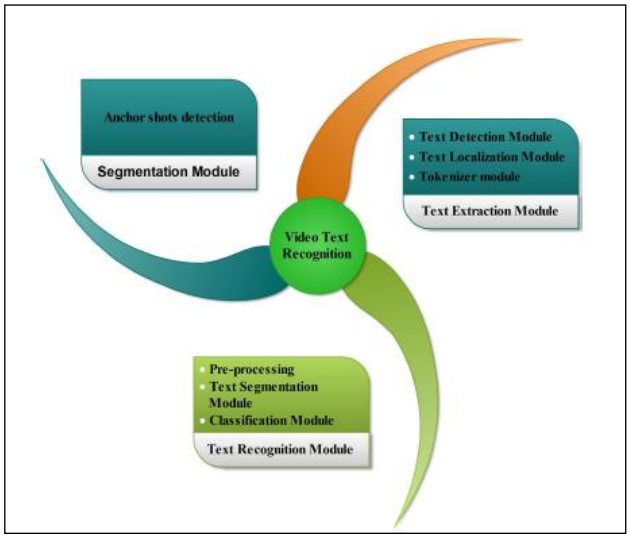


Figure 2. Global Overview of our System

II. SEGMENTATION MODULE

Stories segmentation is an essential step in any work done on video sequences. The story in every news program is the most important semantic unit. Story segmentation is an active research field with a lot of categories of works [16]. For this reason we find a special session concerning video stories segmentation in TRECVID campaigns [15] (TREC Video Retrieval Evaluation) which are the most challenging evaluation for video retrieval in the world. However, the majority of proposed works are based on detecting anchor shots. Indeed, the anchor shot is the only repetitive shot in a news program. The first anchor shot detector dates back to 1995 and it was proposed by Zhang [22]. In these works the anchor shot detection is based on classifying shots according to anchorperson shot models. The drawback of this method is the fact that it is impossible to define a standard model for all channels.

Now, we speak about multimodal anchor shot detection pioneered by Informedia. In this project Yang et al. [21] propose to use high level information (speech, text transcript, and facial information) to classify persons appearing in the news program into three types: anchor, reporter, or person involving in a news event. This method has been proved effective on TRECVID dataset. However, analyzing different video modalities including speech, transcript text, video frames and combining them to extract stories can take a lot of time.

In our system, video sequences are segmented by detecting and classifying faces to find group of anchor shots. This module is based on the assumption that the anchors faces are the only repetitive face throughout the entire program. The features that we extracted from faces will be used to cluster shots. The clustering technique that we used is the Kohonen map. They are well suited for mapping high dimensional vectors (shots) into two dimensional spaces. In fact, the Kohonen map is composed of two layers. The input layer corresponds to the input elements. The output layer corresponds to a set of connected neurons. Every input element is represented by a n-dimensional vector X = (x1, x2, …, xn) and connected to m nodes of the output layer through weights Wij (Figure 3).

After training the Kohonen map, we cluster the nodes of the Kohonen map in order to extract the region (cluster) in which the anchor shot are located. For this reason we used the CMeans classifier. However, the problem is that the C-Means classifier is the fact that it depedns on the number of clusters. It should be provided with the number of clusters to achieve the clustering.

Davies and Bouldin [7], solved this problem by proposing a progressive clustering (2 classes, 3 classes,…, N-max classes). For every clustering they compute a validity index which measures the quality of the clustering. The better classification has the lower of validity index, (Figure 4).

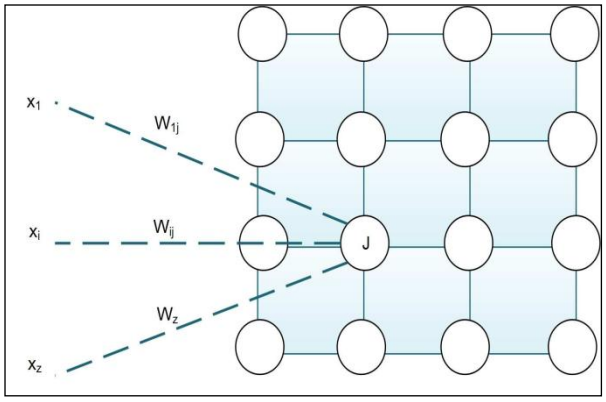


Figure 3. Kohonen Map

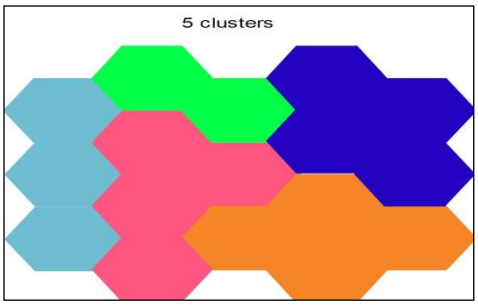


Figure 4. Kohonen map after clustering

After clustering the Kohonen map we have to extract the cluster of anchor shots. For this reason we should consider these two assumptions. First the cluster of anchor shots should not contain less than for shots (faces), since a news story is made up by a minimum of four stories. Second, as the cluster of anchor shots is made by presenter faces only, so the density of this cluster is the smallest one, i.e. the average distance between the faces of the cluster. It is computed as follows:

$$MoyDis(C) = \left(\sum_{i=1}^{n}\sum_{j=i+1}^{n}\left\|X_i - X_j\right\|\right) / (n(n-1)/2) \quad (1)$$

III. TEXT EXTRACTION

Generally, text embedded in video sequences can be classified into two categories: Text that is part of the scene and artificial text. Scene text (such as a sign outside a place of business or placards in front of conference participants) is part of image and usually does not represent information about image content, but artificial text (such as superimposed captions in broadcast news programs and other commercially produced videos) is laid over the image in a later stage. Artificial text is often good key for successful indexing and retrieval of videos. In news broadcast, overlaid text presents the circumstances as places or countries (Iraq, Israel, United States, etc.), gives the names of the interviewed persons or presents an important event (Olympic Games, hostage crisis, etc.). To extract text from stories shots we based on our work proposed in [16].

In this work we use an hybrid approach [14], which combines color and edges to extract text, proceeds as follows: text detection, localization and segmentation. Firstly, for text detection, we apply a new multiple frames integration (MFI) method to minimize the variation of the background of the video frames.

Thereafter, we eliminate from frame columns and rows of pixels which do not contain text. Secondly, we localize text pixels from the remaining rows and column clusters. Every window is represented by two frames. One is the frame of the window filtered along rows and the other one is the frame which is filtered along the columns. For every frame we achieve two operations: First, we realize a transformation from the RGB space to HSV space. Second, we generate, using Sobel filters, an edge picture. For every cluster of these frames, however, we formulate a vector composed of ten features: five of them represent the HSV image and the others represent the edge picture. These features are computed as follows: mean second order moment, third order moment, minimum value of the confidence interval and maximum value of the confidence interval.

For every cluster of these frames, we formulate a vector composed of ten features: five representing the HSV image and five representing the edge picture. These features are computed as follows: mean second order moment, third order moment, minimum value of the confidence interval and maximum value of the confidence interval.

The generated vectors will be presented to a trained back propagation neural network containing ten input nodes, 3 hidden ones and an output node .The training database contains 2000 key frames with the dimension of 320x240. The results of the classifications are two images: an image containing rows considered as text rows and an image containing columns considered as text columns. Finally, we merge results of the two images to generate an image containing zones of text. Once localized, the text in frame will be segmented. The segmentation process is computed as follows:

- Compute the Gray levels image.
- For each pixel in the text area, create a vector composed of two features: the standard deviation and the entropy of the 8 neighborhoods of pixels.
- Run the fuzzy C means clustering algorithm to classify the pixels into “text” cluster and “background” cluster.
- Binarize the text image by marking text pixels in black as shown in (Figure 5).

a) Original frame b) Segmented text frame

Figure 5. Example of text extraction

IV. ARABIC TEXT RECOGNITION

Many successful systems are found for the Latin or Chinese Video Text recognition, but few Arabic Video Text recognition systems are found. This is due to the standard features of Latin or Chinese text that facilitate the recognition process. For example, the Latin character shape do not changes when the character position changes in within a word; characters have only two shapes: capital and small. In opposing to Latin text, Arabic text is cursive and Arabic characters can have four diverse shapes because of their position within the word (Table 1).

Text recognition, even from the detected text lines, remains a challenging problem due to the variety of fonts, colors, presence of complex backgrounds and the short length of the text strings. An optical character recognition methodology was implemented including five successive stages: Pre-processing, Segmentation, Feature extraction, Classification and Post-processing.

A. Morphological Structure of the Arabic Script

The alphabet from the Arabic language has 28 consonants (Table 1) including 15 from one to three points that differentiate between similar characters [Biadsy, 11]. The points and Hamza (ء) called secondary characters (complementary). They are located above the primary character as the "alif" (أ) below as the "Ba" (ب), or in the middle as the "jeem" (ج). There are four characters, which can take the secondary nature Hamzah (ء): alif (إأ), waw (ؤ) kaf (ك) and ya (ئ) .

A distinguishing feature of Arabic writing is the presence of a base-line. The baseline is a horizontal line that runs through the connected portions of text (i.e. where the character's connection segments are located). The baseline has the maximum number of text pixels. Figure 6 demonstrates some of these characteristics on an Arabic sentence *[Mozaffari, 2008] [3]*.

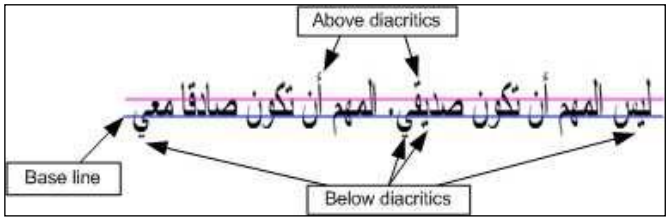


Figure 6. Sample of Arabic text showing some of its characteristics

Arabic script is written and read from right to left. The letters change shape presentation according to their position (at the beginning, middle or end of the word). Each letter can be written Arabic from 2 to 4 different forms *[12]:*

- When the letter is attached to any other: this is the position "isolated";
- When it is attached only to a letter from the front (left) is the position "initial";
- When it is attached only to a letter from behind (right): It is the position "final";
- And when it is attached at both ends at once: this is the position "median".

TABLE I. DIFFERENT FORMS OF ARABIC ALPHABETS

Letter	isolated	Initial	Middle	Final
Alef	ا			ـا
Ba’	ب	بـ	بـ	ـب
Ta’	ت	تـ	تـ	ـت ـة
Tha’	ث	ثـ	ثـ	ـث
Jeem	ج	جـ	جـ	ـج
H’a’	ح	حـ	حـ	ـح
Kha’	خ	خـ	خـ	ـخ
Dal	د			ـد
Thal	ذ			ـذ
Ra’	ر			ـر
Zai	ز			ـز
Seen	س	سـ	سـ	ـس
Sheen	ش	شـ	شـ	ـش
Sad	ص	صـ	صـ	ـص
Dhad	ض	ضـ	ضـ	ـض
Tta	ط	طـ	طـ	ـط
Dha’	ظ	ظـ	ظـ	ـظ
Ain	ع	عـ	عـ	ـع
Gahin	غ	غـ	غـ	ـغ
Fa’	ف	فـ	فـ	ـف
Qaf	ق	قـ	قـ	ـق
Kaf	ك	كـ	كـ	ـك
Lam	ل	لـ	لـ	ـل
Meem	م	مـ	مـ	ـم
Noon	ن	نـ	نـ	ـن
Ha’	ه	هـ	هـ	ـه
Waw	و			ـو
Ya’	ي	يـ	يـ	ـي

B. Existing Arabic OCR

Existing systems for the recognition of the Arabic script is based on the modeling of the three main entities: words, subwords and letters, respectively, and use all three types of approaches: global approach, pseudo-analytical approach and analytical approach *[9]*.

It is through the presentation of a few systems, as is the case for other records, that a comprehensive approach is still limited to a predefined vocabulary words and reduces target for a definite application (cheques, names of cities , and so on.). Systems based on such an approach have always entered into single words. These limits are also laid by Miled and al. *[18]*, and Ben Amara and al. *[1]* as part of their pseudo-analytical approaches based on the rules of concatenation of subwords known for recognizing words (names of cities).

We noticed that in the case of a large or open vocabulary and in the context of recognition of a text for instance, the systems are often based on an analytical approach. Within this framework, several problems were discussed, the most important being the segmentation of the text into words. This problem is more acute for writing Arabic for several reasons: there is an inter-word space and inter-subwords. Another major difficulty to report concerns the establishment of a post-processing because of a lack of vocabulary dictionaries and language tools easily integrated into systems of recognition.

We distinguish post-processing algorithms based on verification and correction by calculating distance editing in a dictionary (combinatorial methods) *[10]*, and the post Treatment based on the statistical structure of the language usually modelled by Hidden Markov Models (statistical methods) *[20] [Khorsheed,03]*.

In contrast to other languages such as English or French whose labels grammatical from a distributional approach characterized by a willingness to set aside any considerations relating to the meaning, the labels come from Arabic of an approach where alongside the formal semantics related to the morphology of the word, without reference to the latter's position in the sentence *[2] [3]*.

This is evidenced by the concepts of patterns and functions that occupy an important place in the grammar of Arabic. For example, the Arabic word “غَلَقَ” (close) is a

verb in the 3rd person of the masculine singular accomplished active, however its form without vowels "غلق" admits grammatical four categories *[4]*:

- Substantive masculine singular "غَلْقٌ" (a closure)
- Word to the 3rd person of the masculine singular accomplished active "غَلَق"or "أغْلَقَ" (he closed)
- Word to the 3rd person of the masculine singular done "غُلِقَ" liabilities (it was closed)
- Word to the imperative 2nd person masculine singular "أغْلِقْ"(closed).

V. Arabic Video Ocr

OCR is one of the most successful applications in the pattern recognition field. It is a common belief that OCR is a solved problem because so many papers and patents have claimed recognition rates as high as 99.99%. Although many commercial OCR systems work well on high quality scanned documents under controlled conditions, they fail in many tasks, such as video OCR, license plate OCR, and sign OCR. Current video OCR is limited to recognizing captions in video images for video indexing, or to identify license plates on vehicles for various applications *[5] [6]*.

The general model of Arabic video text recognition system can be described in terms of five stages: pre-processing, segmentation, feature extraction, classification, post-processing and recognition. Figure 7, illustrates these stages according to their order of occurrence.

The pre-processing stage is a collection of operations that apply some filters that enhance the text extracting from video frames thereby reducing noise and distortion, and consequently get better results from the recognition process. The segmentation stage decomposes the word into characters. The feature extraction stage analyzes a segment and selects a set of features that can be used to uniquely identify the text segment. The classification stage, which is the most the main stage in any OCR system, uses the features extracted in the previous stage to recognize the text segment. Finally, the post-processing stage, which improves the recognition by refining the decisions taken by the previous stage and recognizing words by using context.

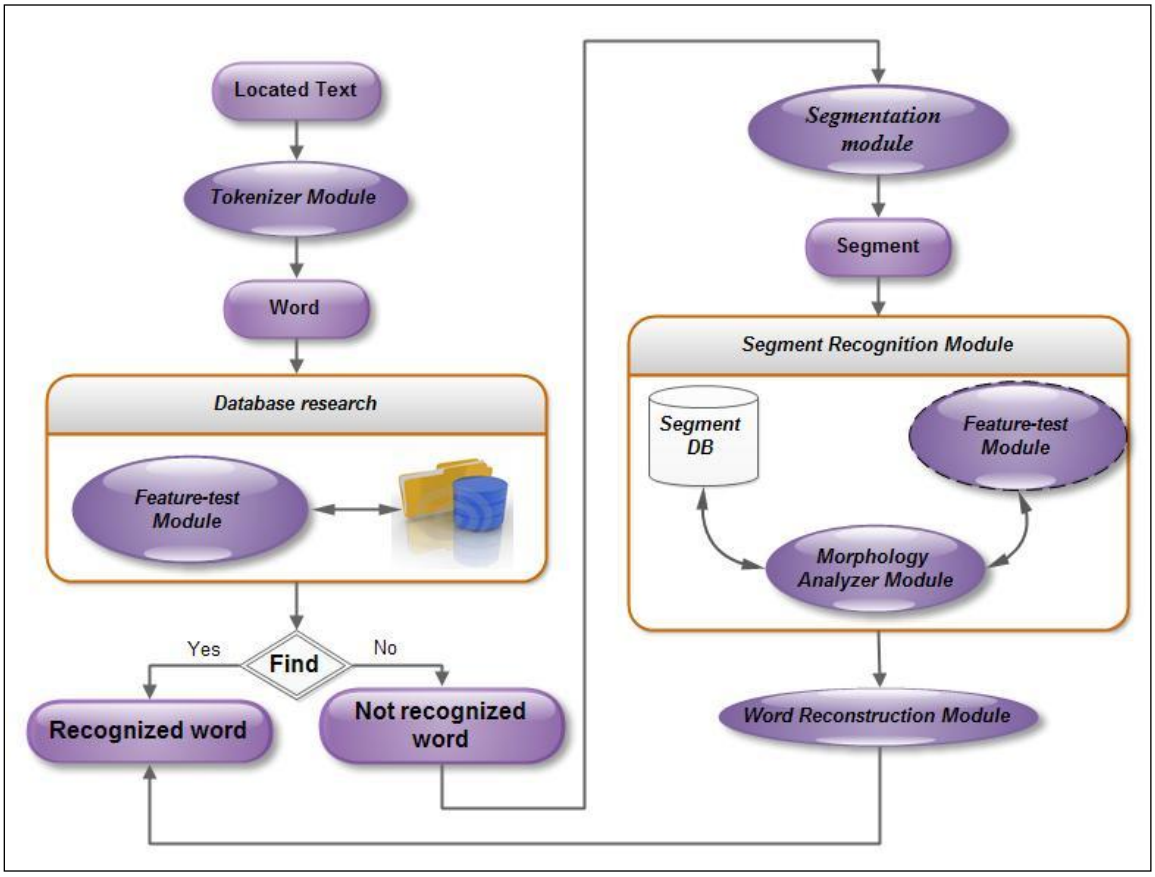


Figure 7. Different stages of proposed OCR system

A. Pre-processing

The recognition accuracy of OCR systems greatly depends on the quality of the input text and noise. The pre-processing stage attempts to compensate for poor quality originals and/or poor quality of Binarization stage after text extraction. This is achieved by reducing both noise and data variations.

1) Text Enhancement

We have chosen the method of Wolf, which uses the contents of all frames of an apparition of the same text to produce an enhanced image. This is done in a robust way, based on statistics calculated on the Gray level of each pixel during the time of onset. To automate this procedure we assume that the text appears only once.

We chose the bilinear interpolation, which calculates the gray level of a pixel as an average of gray levels of its neighbors. The weight of each neighbor is the distance to the pixel calculated:

$$p(x,y) = \frac{1}{\sqrt{(x - ix)^2 + (y - iy)^2}} \quad (2)$$

In different images, text may occur with various widths and heights. To have consistent features through all the text images and to reduce the variance of the size of the characters, the height of a text image is normalized. The height of the text image is scaled to a predefined size (26 pixels).

2) Text Normalization

In different images, text may occur with various widths and heights. To have consistent features through all the text images and to reduce the variance of the size of the characters, the height of a text image is normalized. The height of the text image is scaled to a predefined size (26 pixels).

The size of a character can vary from script to script, which can cause instability of the parameters (descriptors of the character). A natural technique of preprocessing is to bring the characters to the same size. We give in this section an algorithm to normalize the size. We begin by presenting a method of variation of the size of an image. Is an object of size $P \times Q$ we want to reduce the size $M \times N$.

The technique is as follows:

If (i, j) is a pixel in the $M \times N$ final image, the corresponding pixel in the $P \times Q$ source image is:

$$\left(\left[\frac{(i * P)}{N}\right]\right)\left[\frac{(j * Q)}{M}\right] \quad (3)$$

The algorithm of size variation is:

```
1 - Let I and J the source image the final image.
2 - for i = 0 to N-1
    for j = 0 to M-1
J (i, j) = I ([(i * P) / N] [(j * Q) / M])
    End for
End for
```

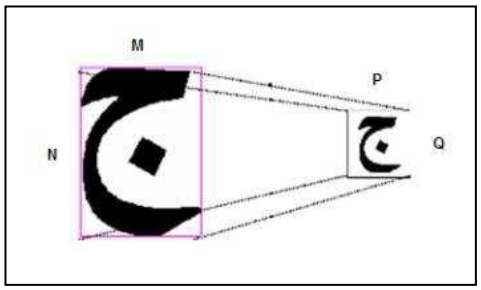


Figure 8. Example of character normalisation

B. Tokenizer module

All components that do not have pixels belonging to the baseline are considered diacritic. The baseline is the one that corresponds to the maximum of the horizontal projection. Information concerning the nature and location of these diacritical marks are registered for use in the future at the stage of recognition. The tokinizer module consists to isolates the words (tokens) of the located text. To build this part we have developed an algorithm that can identify and separate the punctuation marks as well as isolate any extra particles that may be attached to the beginning of the word while they are not part of it. A Database research module is used to check whether the database already contains the word. If the word is not present in the dictionaries, it will be segmented into character using a Segmentation module. Each segment will be recognized using a base of segments.

Detection of diacritics is done in two steps:

- The first step is to perform filtering of connected components based on fairly simple criteria: size of the bounding box, area, vertical layering. The aim is to reject most of diacritics, without rejecting the connected component corresponding to a body of a letter or pseudo-word.
- A second filter allows taking into account the fact that diacritics are either below or above the baseband.

After baseline detection (Figure 9) and diacritic elimination, a segmentation stage is necessary. This segmentation is based on a vertical projection (Figure 10). The segmenter should avoid cases of under-segmentation and/or over-segmentation. Each segment will be recognized using a base of segments.

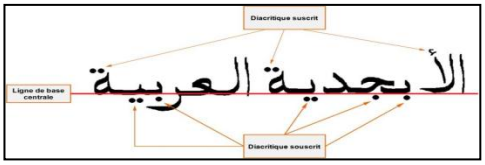


Figure 9. Example of base-line detection

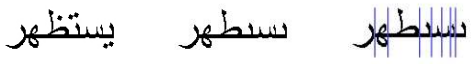


Figure 10. Example of text segmentation

The segmentation of the Arabic word into individual characters is a crucial step in recognizing Arabic text extracting from news sequences. Most of the recognition errors occur from segmentation errors. To minimize error, the segmenter, check before segmented, if there is an intersection pixel between the baseline and the letter to segment (T-junction). This condition gives the following decomposition of letters:

- The letter "ل" at the end of the word is considered as a whole segment.
- The letter "س" will be segmented into three elements.
- The letters "ث, ف, ق ب," will be segmented into two elements, by the segmenter, if they are at the end of the word.
- The letters "ص, ض" will be broken into two segments.

Other letters will be segmented as follows:

- The characters "ث, ف, ب, ث, د, ذ" will be detected by their sizes and positions relative to the subword. Indeed, they are always at the end of the word or subword. Such an error will be corrected by connecting the last segment, with its size, the previous segment.
- For the characters "س, ش, ص, ض", segments have generated a strong resemblance to the characters "ن, ي ب,,ث, ت" at the beginning or middle of a word or subword. A segmentation fault will be corrected by testing diacritics above or below the segment. If the segment contains no diacritics, it is connected to the next segment. The problem here concerns the "1-" character. However, since the last segment of the character has no diacritic, it will be connected directly to the character that precedes it. And since an Arab character and is the most fragmented into three segments, the maximum number of connections should not exceed two connections.
- For the characters " ض,س, ش, ص", the last segment obtained is similar to character" ن". The difference lies in the above diacritic. A segmentation fault will be corrected by testing the diacritic mark. If this diacritic doesn't appear above the segment, (not the character "ن"), it will be connected to the previous segment.

C. Feature extraction

Once an OCR system has an isolated pattern (character or primitive), its next step is to extract the features of the pattern and then classify it. Feature extraction is one of the most difficult and important problems of pattern recognition. Different types and numbers of features may be extracted and used in the recognition stage. The selected set of features should be a small set whose values efficiently discriminate between patterns of different classes, but are similar for patterns within the same class. The feature extraction step is closely related to classification, because the type of features extracted here must match what the classifier expects.

A Feature-test module is used to get the token from the tokenizer module, get some information about it from the morphology analyzer module, and go through several tests one by one until it finds the part of speech of the word. A Morphology analyzer module is called by the feature-test module to analyze all words segments.

To illustrate the sequence of various stages of our system, we present in the following diagram the general synoptic system. The Features chosen to represent each character are:

- Extraction of occlusion: They correspond to the internal contour of the primary plots characters.
- The projection feature: In addition to the horizontal and vertical projection, we make a projection of the diagonal and the slanting diagonal. The projected result will be a vector of size equal to 160.
- Extraction of diacritics marks: diacritics are the secondary parts of characters. These marks above and below the characters, have an important role in the distinction of a few characters that differ only by the number or location of points. The following characters (- اب ت ث -ا) are differed only by the number and location of points.
- The transition features: the number of transition from 0 to 1 of the row, column, diagonal and slanting diagonal. The result is a vector of 160 features for each image of a letter.

D. Classification

Classification in an OCR system is the main decision making stage in which the extracted features of a test set are compared to those of the model set. Based on the features extracted from a pattern, classification attempts to identify the pattern as a member of a certain class. When classifying a pattern, classification often produces a set of hypothesised solutions instead of generating a unique solution.

Supervised classification methods can be used to identify a sample pattern. We used the fuzzy k-nearest neighbour algorithm (k=10). The coefficient of belonging of a new segment x_i to class j is given by the formula:

$$u_{ij} = \frac{\sum_{t=1}^{k} u_{jt} \left(\left\| x_i - x_t \right\|^{m-1/2} \right)}{\sum_{t=1}^{k} \left(\frac{1}{\left\| x_i - x_t \right\|^{m-1/2}} \right)} \quad (4)$$

Where u_{ij} is the coefficient of belonging to class w_j, the tth observation, among the k nearest neighbours of x_i. M is

the variable determines the importance of the contribution of the distance in the calculating of the function of belonging.

E. Post-processing

The recognition rates of character recognition systems are not sufficient, hence is used to improve word recognition rate as opposed to character recognition rate. It is necessary to use context to detect errors and even to correct them. Post-processing is often implemented as a set of techniques that rely on character frequencies, lexicons, and other contextual information. As classification, sometimes, produces a set of possible solutions instead of a unique solution, post-processing is responsible for selecting the right solution using higher level information that is not available to the classifier. Post-processing also uses that higher level information to check the correctness of the solutions returned by the classifier. The post-processing operations are spell checking and correction. Spell checking can be as simple as looking up words in a lexicon.

In this stage a Word reconstruction module is used to gives words assumptions from segments previously recognized. The validation of alternative assumptions words is a user interactive job.

The post-processing stage, which is the final stage, improves recognition by refining the decisions taken by the previous stage and recognises words by using context. It is ultimately responsible for outputting the best solution, and is often implemented as a set of techniques that rely on character frequencies, lexicons, and other context information.

VI. Experiments

To validate our approach, we have used a varied database composed of news sequences extracted from different Arabic channels. Concerned channels are Tunisia 1 TV (tunisiatv.com) as generalist channel presenting news at 13H and 20H. We have also tried our approach with news video extracted from Aljazeera (Aljazeera.net) and AlArabiya (alarabiya.net) which are specialist channels presenting news continually.

A. Experimental results of extraction text

To experiment our previous work we have based on the number of identified correct text contours, we find the following recall and precision rate to evaluate the extraction text method. Figure 11 shows the results of this experiment.

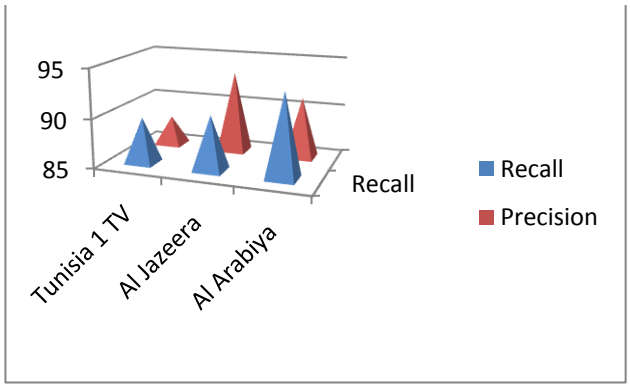


Figure 11. Evaluation results of textual localization

We notice that Al Jazzeera and Al Arabiya channels present the best rate of recall and precision because they present the best quality of graphic text. In fact, our approach is more robust to various font size, font styles, contrast levels and background complexities because it uses both of colour features and edges features to differentiate text pixels from background pixels, besides it's based on a neural network trained on different types of text styles.

B. Experimental results of Recognition text

In order to investigate the effectiveness of our recognition sub-system in recognizing Arabic Text extracted from various news sequences, a series of tests were performed using fuzzy KNN classifier with K=3, K=5 and K=10. Figure 12 shows the recognition results.

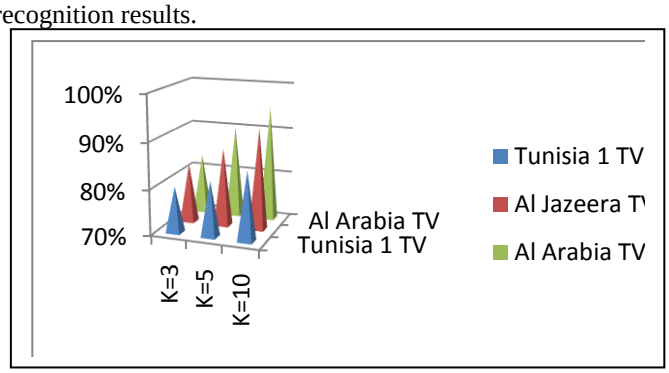


Figure 12. Evaluation results of text recognition

We note that the rate of recognition of texts extracted from Al-Arabiya and Al-Jazeera TVs are better than those extracted from Tunisia 1 TV. This is explained by the fact that the text extract from Al-Arabiya and Al-Jazeera TVs is clearer and more readable. We find it is difficult to recognize small size characters that appear frequently in news videos. The best recognition rate is in all cases for K=10.

This recognition system was also tested on the database "Houda"[17]as part of "ICDAR 2009 Handwritten Farsi/Arabic Character Recognition Competition" [19]. Hoda database which contains handwritten digits is presented by Khosravi and Kabir in 2007 [Khosravi, 2007]. Binary images of 102,352 digits were extracted from about 12,000 registration forms. These forms were scanned at 200 dpi with a high speed scanner. Hoda database was partitioned into train (60,000 samples) and test (20,000 samples) subsets.

VII. Conclusion

In this paper, we propose an efficient method to deal with background complexity in Arabic news video text recognition by efficiently integrating multiple frame information. Proposed system includes several tasks such as Text Detection; Localization; Segmentation; and Recognition. By using this system we can produce quite clear text for Arabic text recognition. The extraction rate has been increased about 91% and 84% for text recognition. These methods can also be adopted by any other type of Video OCR systems to increase recognition rate.

Used OCR techniques do not perform very well when the extracted characters have a too low resolution. The development of new OCR techniques to recognize low

resolution characters is still necessary. Other aspect is computation reduction for mobile image text recognition applications. Most mobile devices, such as mobile phones, have less computation power and less memory resources than a desktop computer. In order to build a video text extraction application on these devices, the algorithms proposed in this paper need to be optimized or even modified to reduce the computation cost.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program.

REFERENCES

[1] N. E. Ben Amara, A. Belaïd, N. Ellouze, "Utilisation des modèles markoviens en reconnaissance de l'écriture arabe : Etat de l'art, CIFED'2000, 181-191 (2000).

[2] M. Ben Halima, H. Karray and A. M. Alimi, Arabic Text Recognition in Video Sequences, International Journal of Computational Linguistics Research, Volume 1, Issue 2, June, 2010, Pages 72-80.

[3] M. Ben Halima, H. Karray and A. M. Alimi, A Comprehensive Method for Arabic Video Text Detection, Localization, Extraction and Recognition, The 2010 Pacific-Rim Conference on Multimedia (PCM2010), September 21-24, Shanghai, China. Springer-Verlag Berlin, Heidelberg ©2010. Lecture Notes in Computer Science, 2011, Volume 6298/2011, pp. 648-659.

[4] M. Ben Halima, H. Karray and A. M. Alimi, "AVOCR: Arabic Video OCR", The International Symposium on Image/Video Communications over fixed and mobile networks (ISIVC 2010), September 30, October 1-2, 2010, Rabat, Morocco.

[5] Y. Chang, Y. Zhang, S. Vogel, J. Yang, \Enhancing Image-based Arabic Document Translation Using a Noisy Channel Correction Model," In Proceedings of MT Summit XI, Copenhagen, Denmark, Sep. 10-14, 2007.

[6] Y. Chang, D. Chen, Y. Zhang, J. Yang, "An Image-based Automatic Arabic Translation System," Pattern Recognition, Vol. 42, pp. 2127-2134, 2009.

[7] Davies D. L. and Bouldin D. W., "A cluster separation measure", In IEEE Transaction Pattern Analysis Machine Intelligence, vol. PAMI-1,(1997), pp 224–227.

[8] C. Garcia, X. Apostolidis, Text detection and segmentation in complex color images, in International Conference on Acoustics, Speech and Signal Processing, 2000 ,pp. 2326-2329.

[9] L. Hamami and D. Berkani, "Recognition System for Printed Multi-Font and Multi-Size Arabic Characters", The Arabian Journal for Science and Engineering, Volume 27, Number 1B, 57-72 (2002).

[10] L. Hamami and D. Berkani, "Recognition System for Printed Multi-Font and Multi-Size Arabic Characters", The Arabian Journal for Science and Engineering, Volume 27, Number 1B, 57-72 (2002).

[11] K. Jung, K.I. Kim, and A.K. Jain. Text information extraction in images and video: a survey. Pattern Recognition, pp. 977-997, May 2004.

[12] S. Kanoun, A. Alimi, Y. Lecourtier, Affixal Approach for Arabic Decomposable Vocabulary Recognition: A Validation on Printed Word in Only One Font, ICDAR'2005, pp. 1025–1029.

[13] H. Karray, A. M. Alimi, Detection and Extraction of the Text in a video sequence, in Proc. IEEE 12 International Conference on Electronics, Circuits and Systems 2005 (ICECS 2005), vol. 2, pp. 474–478

[14] Karray H. , Ellouze M, Alimi A.M.,"KKQ: K-frames and K-Words Extract for Quick News Story",International Journal of Information and Communication Technology, Vol.1, / ISSN: 0973-5836 /,pp 69-76, June 2008.

[15] H. Karray, A. Wali, N. Elleuch, A. Ammar, M. Ellouze, I Feki, A. M. Alimi, High-level Features Extraction and Video Search, TRECVID2008, http://www-nlpir.nist.gov/projects/tv2008/tv2008.html

[16] Kherallah M., Karray H., Ellouze M., Alimi A.M., Toward an Interactive Device for Quick News Story Browsing, icpr 2008, pp 1-4.

[17] H.Khosravi and E.Kabir, Introducing a very large dataset of handwritten Farsi digits and a study on their varieties, Pattern Recognition Letters 28 (2007) 1133–1141.

[18] H. Miled, N. Ben Amara, "Planar Markov Modeling for Arabic Writing Recognition: Advancement State", Proceeding. of ICDAR'01, 69-73 (2001).

[19] Saeed Mozaffari and Hadi Soltanizadeh, ICDAR 2009 Handwritten Farsi/Arabic Character Recognition Competition, ICDAR 2011, pp. 1413-1417.

[20] M. Pechwitz and V. Maergner, "HMM Based Approach for Handwritten Arabic Word Recognition Using the IFN/ENIT- Database", Proceedings of ICDAR, 890-894 (2003).

[21] L. Yang, L. Hong, X. Xiangyang, T. Yap-Peng, Effective video text detection using line features, ICARCV 2004, pp 1528-1532, December 2004.

[22] H.J. Zhang, C. Y. S. Low, W. Smoliar, D. Zhong, "Video parsing, Retrieval and Browsing: An integrated and Content-Based Solution", In Proceedings of ACM Multimedia (1995), pp45-54.

Secured Wireless Communication using Fuzzy Logic based High Speed Public-Key Cryptography (FLHSPKC)

Arindam Sarkar
Department of Computer Science & Engineering
University of Kalyani
Kalyani-741235, Nadia, West Bengal, India.

J. K. Mandal
Department of Computer Science & Engineering
University of Kalyani
Kalyani-741235, Nadia, West Bengal, India.

***Abstract*— In this paper secured wireless communication using fuzzy logic based high speed public-key cryptography (FLHSPKC) has been proposed by satisfying the major issues likes computational safety, power management and restricted usage of memory in wireless communication. Wireless Sensor Network (WSN) has several major constraints likes' inadequate source of energy, restricted computational potentiality and limited memory. Though conventional Elliptic Curve Cryptography (ECC) which is a sort of public-key cryptography used in wireless communication provides equivalent level of security like other existing public–key algorithm using smaller parameters than other but this traditional ECC does not take care of all these major limitations in WSN. In conventional ECC consider Elliptic curve point *p*, an arbitrary integer *k* and modulus *m*, ECC carry out scalar multiplication *kP* mod m, which takes about 80% of key computation time on WSN. In this paper proposed FLHSPKC scheme provides some novel strategy including novel soft computing based strategy to speed up scalar multiplication in conventional ECC and which in turn takes shorter computational time and also satisfies power consumption restraint, limited usage of memory without hampering the security level. Performance analysis of the different strategies under FLHSPKC scheme and comparison study with existing conventional ECC methods has been done.**

Keywords- Soft computing; Wireless Communication; High Speed; ECC.

I. INTRODUCTION

Now-a-days a number of public-key cryptographic algorithms are obtainable for securing data like authentication, non-repudiation, integrity and confidentiality [7, 8, 16]. Among these algorithms a few of them are faster and some of the other algorithms are slower. WSN consist of small nodes that sense their surroundings, process data and communicate through wireless connection. In wired data networks node rely on pre deployed trusted server assist to set up belief connection but in case of WSN, because of limited memory CPU power , and energy of sensors their trusted authorities do not exist [17].

Due to the restraint in the bandwidth, computational potency, power availability or storage in mobile devices, the PKC-based remote authentication schemes are not appropriate for mobile devices.

For these reasons WSN needs faster public-key cryptographic algorithm which satisfied all these constraints [18, 19, 20, 21]. Neal Koblitz and victor Miller was proposed Elliptic Curve Cryptography (ECC) [2]. The advantages of ECC over RSA, DSA, ElGamal [3], Rabin, Diffie-Hellman Key exchange [1] are that it provides efficient security using shorter key and offers

1) *minimum space complexity for key storage.*
2) *less energy cost for performing arithmetic operations.*
3) *minimum time complexity for transmission of keys.*

These features make ECC best substitute to offer security in WSN [15]. ECC offers a popular solution to the problem of implementing public key cryptography on mobile computing devices. The security of RSA, the most popular algorithm in other fields is based on the hardness of integer factorization. ECC is based on the Elliptic Curve Discrete Logarithm Problem (ECDLP) which is the best algorithm for solving ECC takes fully exponential time [15].

Consider base point P, arbitrary integer k, scalar multiplication will be Q = kP. Using ECDLP to find k from P and Q are a hard problem. Whereas the best algorithm for solving RSA, DSA takes sub exponential amount of time [12].

ECC keys are shorter than their RSA analogues, while achieving the similar security level. A 160-bit ECC key is approximately corresponding to a 1024-bit RSA key. So ECC based system is normally more competent and utilize less resources than RSA and hence ECC has appeared as a promising choice to traditional public key techniques on WSNs, because of its lesser processing and storage requirements. ECC is faster than RSA for decryption, but slower than RSA for encryption.

In ECC-based authentication algorithm, Elliptic Curve scalar multiplication is core operation, but this operation is the most time consuming operation. This operation takes 80% of executing time.

Hardware implementation of ECC involves tree-layer hierarchical strategy namely finite field arithmetic, point arithmetic and scalar multiplication as shown in figure 2.

There are many attempts to implement this 3-tier in order to obtain swift computation, reduce power consumption and reduce storage space.

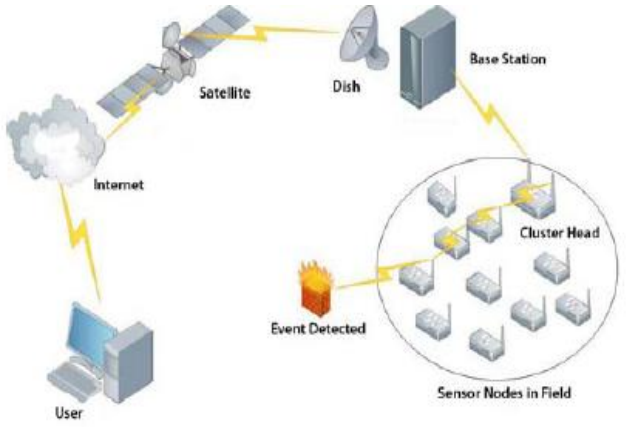


Figure 1. Structural design of WSN

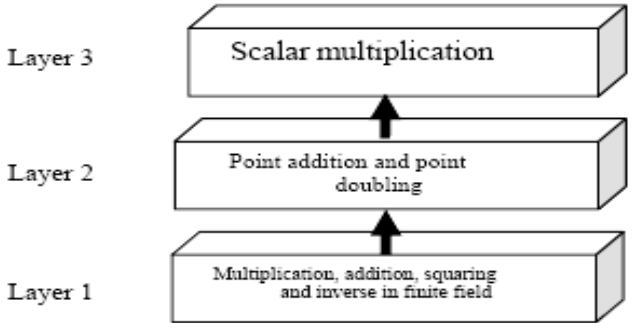


Figure 2. 3-tier architecture of ECC scalar multiplication

In this paper for implementing ECC over binary fields, improving its performance and viability an optimized technique has been proposed. The most important operations in any ECC based techniques such as key exchange or encryption is the scalar multiplication. This point scalar multiplication is achieved by repeated point addition and doubling.

In this paper FLHSPKC scheme has been proposed for scalar multiplication which outstandingly improves the computational competence of scalar multiplication. In proposed FLHSPKC scheme 4 techniques has been proposed to extend the performance of classical ECC by providing following features and comparisons among these newly proposed scheme has also been done.

- Requirement of less amount of memory during execution
- Does not store individual operation at some point that means it do not store pre computed values but it keeps full power computation.
- It is as fast as the traditional ECC.

II. Related Work

Gura *et al.* [14] showed that scalar multiplication which is the operation of multiplying point *p* on an elliptic curve *E* defined over a field *GF*(*p*) with positive integer *k* which involves point addition and point doubling, spent 80% of execution time. Also the efficiency of operation *kP* is depends on the type of coordinate system used for point *p* on the Elliptic Curve and the algorithm used for recoding of integer *k* in scalar multiplication. An ample amount of analysis on binary fields and Elliptic Curve arithmetic for the NIST recommended elliptic curves was done in [4]. A dedicated implementation for the field GF (2^{155}) was done in [10]. More point multiplication algorithms can be found in [5] and [6]. Shantz [11] presented an efficient technique to calculate modular division, which is an vital arithmetic operation in ECC and other cryptographic system. Cohen et al [9] analyzed the impact of coordinate system in ECC implementation. They measured the performance of point Addition (PADD) and Point Doubling (PDBL) of different coordinate system. Malan et.al [13] implemented an ECC system using polynomial basis over binary field GF(2^m).

So, aim of this paper is to propose a novel technique which can reduce the time required in scalar multiplication.

III. Introduction to Elliptic Curve Cryptography

An elliptic curve E over $GF(P)$ can be defined by

$$y^2 = x^3 + ax + b \text{ where } a,b \in GF(p) \quad (1)$$

$$4a^3 + 27b^2 \neq 0 \quad (2)$$

in the $GF(P)$. The point *(x, y)* on the curve satisfies the above equation and the point at infinity denoted by ∞ is said to be on the curve.

For example, $y^2 = x^3 + 2x + 5$ (3)

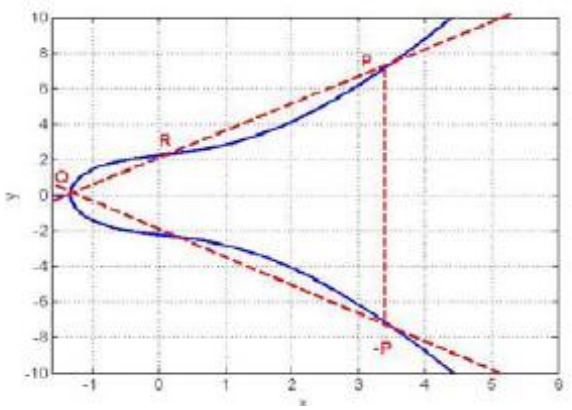


Figure 3. Elliptic Curve of equation (3)

$$y^2 = x^3 - 2x + 1 \quad (4)$$

The Elliptic Curve group operation is closed so that the addition of any two points is again a point of Elliptic Curve. Identity element of the group is point at infinity i.e. O and stratifies $P + O = O + P = P$.

For *P(x, y)* it can be define that $-P = (-x, y)$ as the unique inverse of *P* in the group satisfying the property $P + (-P) = (-P) + P = O$.

A multiplication of a point *P* with an integer *k* can be expressed as the multiplication addition of one and the same point *k*-times.

$$Q = P + P + P + \ldots + P = k.P \quad (5)$$

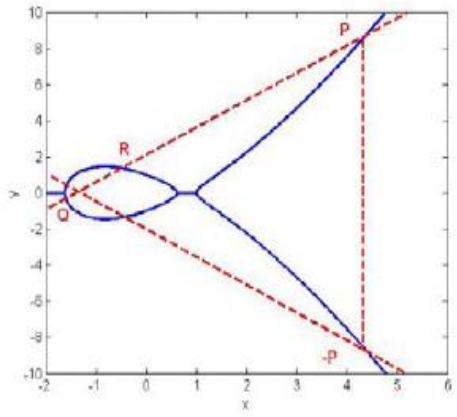


Figure 4. Elliptic Curve of equation (4)

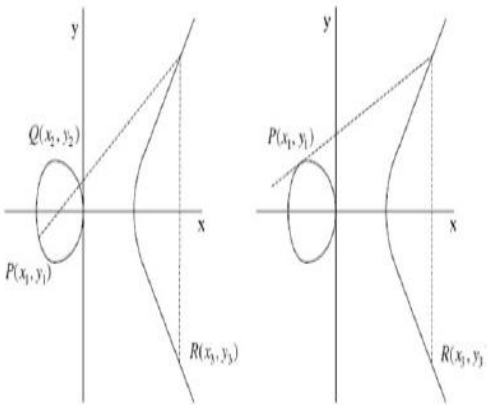
The resultant product *Q* is another point on the Elliptic curve. Given an Elliptic Curve *C* over finite field F_2^p, a point *P* and a product *Q* the problem is to find a *k.N* that holds *Q=k.P*. This problem is known as Elliptic Curve Discrete Logarithmic Problem (ECDLP) which is quite hard to solve. For example with a finite fields with F_2^p, 2^p elements takes about $O(2^{P/2})$operations to find *k which* is exponential amount of time. Thus the ECC depends on the complexity of the ECDLP, that is given points P and Q in the group, to find the number k such that Q=k.P. If there are two points on the curve namely, $P(x_1, y_1)$, $Q(x_2, y_2)$ and their sum is given by point $R(x_3, y_3)$ the algebraic formulas for point addition and point doubling are given by following equations:

$$x_3 = \lambda^2 - x_1 - x_2 \quad (6)$$

$$y_3 = \lambda(x_1 - x_3) - y_1 \quad (7)$$

$$\lambda = \frac{(y_2 - y_1)}{(x_2 - x_1)}, if \quad P \neq Q \quad (8)$$

$$\lambda = \frac{(3x^2 + a)}{2y_1}, if \quad P = Q \quad (9)$$



(a) Addition: P + Q = R (b) Doubling: P+P = R

Figure 5. Elliptic Curve point addition and point doubling operations

IV. Proposed Flhspkc Technique

For reducing computational complexity of scalar multiplication proposed FLHSPKC techniques provides some novel schemes given in the following subsections.

A. Hamming Weight Reduction Strategy

The Hamming weight of a binary string is the number of symbols that are diverse from the "0" symbol. So in general Hamming weight is the number of "1" bits presence in the binary sequence. PADD is more computationally expensive than PDBL. Now, point multiplication can be decomposed to series of PADD and PDBL, because of computational complexness PDBL should be use more than PADD for point multiplication computational purpose. As an example consider $k= 511= (111111111)_2$. Now for any point *P*, *511P* requires 8 point additions. However, if we write $(111111111)_2 = (100000000)_2 - 1$. Then in decimal $511P = 512P - P = (100000000)_2P - P$ requires only one addition. Again point subtraction can be replaced by PADD because the inverse of an affine point *P (x, y)* is *-P (x, -y).* Thus this proposed strategy can implement in point multiplication to attain faster computation.

B. Pairing Strategy

Here pairing strategy for scalar multiplication in ECC has been used for computing *kP* for any $k \in Z^+$ (set of positive integers). Let *P* be a point on EC and let *k=3277* then *3277P* is equivalent to $(110011001101)_2P$. Consider pairing size from 2 to 12; but it can be of any size. For a pairing size *PSIZE*, there are $(2^{PSIZE-1}-1)$ pre computations are needed i.e. if *PSIZE =3* then there are, $(2^{3-1}-1) = 3$ pre computations, namely *3P*, *5P* and *7P*. So increasing of pairing size also increase the number of pre computation and number of additions and doubling operations decreases. The numbers of PADD and PDBL for 3277P for different pairing size are as follows.

- Consider the pairing size *PSIZE = 2* and *k= 763* then $763 = (1011111011)_2$

  No of pre computations = $2^{PSIZE-1}-1 = 2^2 - 1 = [3]\ P$

  So, $763 = \underline{10}\ \underline{11}\ \underline{11}\ \underline{10}\ \underline{11}$

  The intermediate values of *Q* are *P*, *2P 4P*, *8P*, *11P*, *22P*, *44P*, *47P*, *94P*, *95P*, *190P*, *380P*, *760P*, *763P*

  Computational cost = 9 doublings, 4 additions, and 3 pre computation.

- Consider the pairing size *PSIZE = 3* and *k= 763* then $763 = (1011111011)_2$

  No of pre computations = $2^{PSIZE-1}-1 = 2^3 - 1 = [7]\ P$

  The intermediate values of *Q* are *5P*, *10P*, *20P*, *40P*, *47P*, *94P*, *188P*, *376P*, *381P*, *762P*, *763P*

  Computational cost = 7 doublings, 3 additions, and 7 pre computations.

- Consider the pairing size *PSIZE = 4* and *k= 763* then $763 = (1011111011)_2$

No of pre computations = $2^{PSIZE-1}-1 = 2^4 - 1 = [15]\ P$

The intermediate values of *Q* are *11P, 22P, 44P, 88P, 95P, 190P, 380P, 760P, 763P*

Computational cost = 6 doublings, 2 additions, and 15 pre computations.

- Consider the pairing size *PSIZE = 5* and *k= 763* then $763 = (1011111011)_2$

  No of pre computations = $2^{PSIZE-1}-1 = 2^5 - 1 = [31]\ P$

  The intermediate values of *Q* are *23P, 46P, 92P, 184P, 368P, 736P, 763P*

  Computational cost = 5 doublings, 1 additions, and 31 pre computations.

- Consider the pairing size *PSIZE = 6* and *k= 763* then $763 = (1011111011)_2$

  No of pre computations = $2^{PSIZE-1}-1 = 2^6 - 1 = [63]\ P$

  The intermediate values of *Q* are *47P, 94P, 188P, 376P, 752P, 763P*

  Computational cost = 4 doublings, 1 additions, and 63 pre computations.

- Consider the pairing size *PSIZE = 7* and *k= 763* then $763 = (1011111011)_2$

  No of pre computations = $2^{PSIZE-1}-1 = 2^7 - 1 = [127]\ P$

  The intermediate values of *Q* are

  95P, 190P, 380P, 760P, 763P

  Computational cost = 3 doublings, 1 additions, and 127 pre computations.

- Consider the pairing size *PSIZE = 8* and *k= 763* then $763 = (1011111011)_2$

  No of pre computations = $2^{PSIZE-1}-1 = 2^8 - 1 = [255]\ P$

  The intermediate values of *Q* are *95P, 190P, 380P, 760P, 763P*

  Computational cost = 3 doublings, 1 additions, and 255 pre computations.

- Consider the pairing size *PSIZE = 9* and *k= 763* then $763 = (1011111011)_2$

  No of pre computations = $2^{PSIZE-1}-1 = 2^9 - 1 = [511]\ P$

  The intermediate values of *Q* are *381P, 762P, 763P*

  Computational cost = 1 doublings, 1 additions, and 511 pre computations.

- Consider the pairing size *PSIZE = 10* and *k= 763* then $763 = (1011111011)_2$
  No of pre computations = $2^{PSIZE-1}-1 = 2^{10} - 1 = [1023]\ P$
  The intermediate values of *Q* are *763P*
  Computational cost = 0 doublings, 0 additions, and 1023 pre computations.

C. 1's Complement based Arithmatic Strategy

In FLHSPKC scheme there exist another strategy based on 1's complement arithmetic. The formula used for calculating 1's complement is

$$Comp = (2^{\text{no. of bits}} - 1) - N \quad (9)$$

Where *Comp* = 1's complement of the binary number, *N* = respective binary number and

no. of bits = number of bits presence in binary form of *N*. To reduce number of intermediate operations of multiplication, squaring and inverse calculations used in ECC minimal non-zero bits in positive integer scalar are very important Now, equation (9) can be written as

$$N = (2^{no.\ of\ bits} - 1 - Comp) \quad (10)$$

For example, if *N =2046* then binary representation of *N* will be $N= (11111111110)_2$. Now calculate 1's complement of *N* i.e. $Comp = (00000000001)_2$. Number of bits present in binary representation of *N* is 11. After putting the respective values of *N, Comp and no. of bits* in equation (10) we get $2046 = (2^{11} - 00000000001 - 1)$, this can be reduced as $2046 = (100000000000)_2 - (00000000001)_2 - 1$

So we have, $2046 = (100000000000)_2 - (2^1 - 0 - 1) - 1$ i.e.
$2046 = 100000000000 - 1 - 1 = 2048 - 1 - 1$.

The Hamming weight of *N* has reduced from 10 to 1 which will save 9 elliptic curve addition operations. For providing more optimized result above discussed 1's complement subtraction method is combined with sliding window strategy.

Now apply this proposed 1's complement arithmatic strategy to the same number 763 to show the effectiveness of algorithm with previously discussed proposed pairing strategy with *PSIZE=3*.

$763 = (1011111011)_2$

Using equation (10) we can write

$763 = (10000000000)_2 - (0100000100)_2 - 1$

The intermediate values of Q are:

3P, 6P, 12P, 24P, 48P, 96P, 192P, 384P, 768P, 763P

Hence the Computational Cost = 8 doublings, 1 addition and 3 pre computations.

With *PSIZE*= 3, in pairing scheme for the same examples the intermediate values of *Q* are *5P, 10P, 20P, 40P, 47P, 94P, 188P, 376P, 381P, 762P, 763P and* Computational cost = 7 doublings, 3 additions, and 3 pre computations. So using 1's complement arithmetic scheme the computational cost has been reduced from 3 additions as in binary method to only 1 addition. The number of pre computations remained same. This can be proved for different *PSIZE*.

D. Soft Computing based Arithmatic Strategy

In this proposed soft computing based scheme fuzzy logic approach is used to perform recurring subtraction instead of performing division. For that some random multiples of m is used. This operation has certain fuzziness because the operation used is not exact but it involves some random multiples of m from the value to be reduced modulo m. The example of proposed fuzzy logic based scheme is shown in figure 6 and traditional binary multiplication of 2 binary integers is shown in figure 7. Let X=26 and Y=24 with m=17 then value of $X * Y \bmod m$ will be 12. Since, fuzzy

based approach mainly focuses on repeated subtraction instead of division for modulo reduction m. The process is shown in figure 6. The modulus m=17, a multiple of m is consider for speed of this process i.e. m.t=17*5=85.

- Whenever from number Y, " 0" will be encountered then bit position weight from Y for corresponding "0" (find out bit position weight using 8-4-2-1 rules) will be multiplied by m.t value i.e. (Bit position weight x m.t) and this value will be subtracted.

- Whenever from number Y, " 1" will be encountered then bit position weight from Y for corresponding "1" (find out bit position weight using 8-4-2-1 rules) will be multiplied by X value i.e. (Bit position weight x X) and this value will be added.

$X = (26)_{10} = (11010)_2$ and $Y = (24)_{10} = (11000)_2$

$(26)_{10}$ $(11010)_2$
$\times (24)_{10}$ $(11000)_2$
-1×85
-2×85
-4×85
$+8 \times 26$
$+16 \times 26$
29

$(29)_{10} \bmod (17)_{10} = (12)_{10}$

Figure 6. Soft Computing based modular multiplication

$X = (26)_{10} = (11010)_2$ and $Y = (24)_{10} = (11000)_2$

11010
×11000
00000
00000×
00000×
11010×
11010×
1001110000 = $(624)_2$

$(1001110000)_2 \bmod (1001)_2 = (110)_2 = (12)_{10}$

Figure 7. Traditional modular multiplication

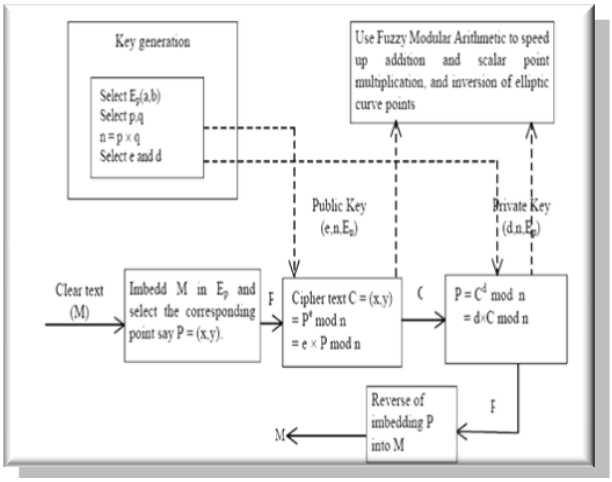


Figure 8. Soft Computing based Arithmatic Model

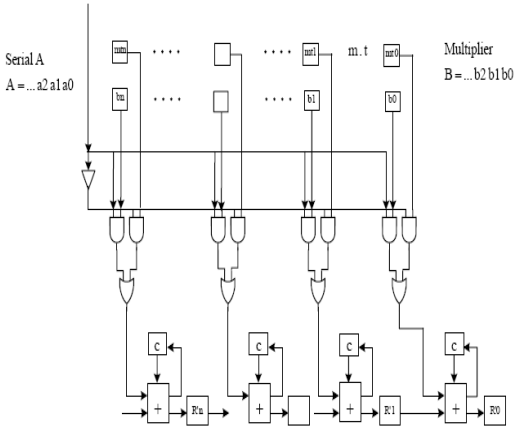


Figure 9. Array Multiplier for Soft Computing based Arithmatic

V. PROPOSED SOFT COMPUTING BASED CONTROLLER FOR ECC

There is a tradeoff between the computational cost and the window size (*PSIZE* in proposed pairing scheme). Now this tradeoff is depends on balance between computing cost (or the RAM cost) and the pre computing (or the ROM cost) of the node. From this discussion it is noticed that the variety of wireless network working states will make this control complex and calculations could be relatively more expensive. Therefore, FLHSPKC ensures the optimum window size (*PSIZE* in proposed pairing scheme) is obtained by tradeoff between pre computation and computation cost with the help of fuzzy dynamic control system, to provide dynamic control. The goal of fuzzy decision problem is to maximize the minimum value of the membership functions for optimization purpose. Representation of multiobjective programming problem with the help of fuzzy optimization model is discussed follows.

$$Max: \min\{\mu_a(\phi)\} \& \min\{\mu_b(\varphi)\} \quad \forall a \in A \& \forall b \in B \qquad (11)$$

Such that $X_b \le Y_l \qquad \forall b \in B \qquad (12)$

$$\sum_{r \in R_p} z_{ra} = 1 \qquad \forall p \in P \& \forall a \in A \qquad (13)$$

$$z_{ra} = 0 \text{ or } 1 \qquad \forall r \in R \& \forall a \in A \qquad (14)$$

Objective of the above equation is to maximize the minimum membership function of delays ϕ and φ denotes the deviation value between recommend value and measured value.

Figure 8 shows a fuzzy control system with 3 inputs fuzzy controller. They are as follows.

- *Storage Room: it is the first input having* 3 states, namely (i) Min (ii) Intermediate (iii) Max
- *PreComputing:* The second input is pre-computing working load (*PreComputing*) in one of three states, namely (i) Min (ii) Intermediate (iii) Max
- *Doubling* : The third input is *Doubling*, expressing how much working load for the calculation "doubling" which has three states, namely (i) Min (ii) Intermediate (iii) Max

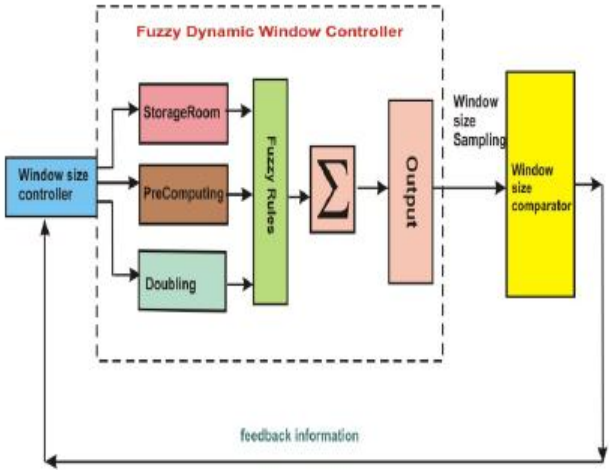


Figure 10. Proposed Soft Computing based Controller for ECC

The output is one, i.e.

- *WindowSize:* To express the next window size should be moved in which way, which has three states for the window sizes, namely (i) down (ii) stay (iii) up.

There are 26 *Fuzzy Rules* listed as shown in Table I where weights are unit.

- Mi=Min
- I=Intermediate
- Mx= Max
- D= down
- S= stay
- U=up

TABLE I. SHOWS 26 FUZZY RULES

StorgeRoom	PreComputing	Doubling	WindSize
Mi	Mi	Mi	U
Mi	Mi	I	U
Mi	Mi	Mx	S
Mi	I	Mi	U
Mi	I	I	U
Mi	I	Mx	S
Mi	Mx	L	U
Mi	Mx	I	S
Mi	Mx	Mx	S
I	Mi	Mi	U
I	Mi	A	U
I	Mi	Mx	S
I	I	L	U
I	I	I	S
I	I	Mx	D
I	Mx	I	S
I	Mx	Mx	S
Mx	Mi	Mi	S
Mx	Mi	I	S
Mx	Mi	Mx	D
Mx	I	Mi	S
Mx	I	I	S
Mx	I	Mx	D
Mx	Mx	Mi	D
Mx	Mx	I	D
Mx	Mx	Mx	D

From the above table I it is observed that among 26 fuzzy rules only 9 fuzzy rules will be dominated the whole fuzzy control system. Considering only major 9 fuzzy rules quality of service (QOS) can be improved by decreasing the latency time of the system. These major 9 fuzzy rules are as follows:

1 *If (PreComputing is Min) and (Doubling is Min) then (WindowSize is Up)*

2 *If (PreComputing is Min) and (Doubling is Intermediate) then (WindowSize is Up)*

3 *If (PreComputing is Min) and (Doubling is Max) then (WindowSize is stay)*

4 *If (PreComputing is Intermediate) and (Doubling is Min) then (WindowSize is Up)*

5 *If (PreComputing is Intermediate) and (Doubling is intermediate) then (WindowSize is Up)*

6 *If (PreComputing is intermediate) and (Doubling is Max) then (WindowSize is stay)*

7 *If (PreComputing is Max) and (Doubling is Min) then (WindowSize is Up)*

8 *If (PreComputing is Max) and (Doubling is intermediate) then (WindowSize is stay)*

9 *If (PreComputing is Max) and (Doubling is Max) then (WindowSize is stay)*

In each above fuzzy conditions the value within a bracket denotes the weight number in unit. With 3 inputs StorageRoom, PreComputing and Doubling and 1 output i.e. WindowSize Mamdani composition rule base is used in each 9 fuzzy rules.

VI. RESULTS AND DISCUSSION

In this section results of each proposed novel strategy under FLHSPKC scheme has been presented in the following.

TABLE II.

PAIRING SIZE VS NO OF DOUBLINGS, ADDITIONS AND PRE COMPUTATIONS IN PAIRING STRATEGY

PSIZE	Number of Doublings	Number of Additions	Number of Pre-computation
2	9	4	3
3	7	3	7
4	6	2	15
5	5	1	31
6	4	1	63
7	3	1	127
8	3	1	255
9	1	1	511
10	0	0	1023

Figure 11 shows the tradeoff between window size and the computational costs in Pairing strategy.

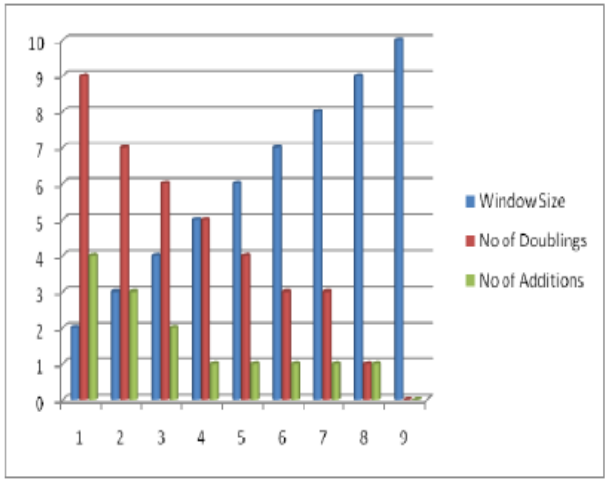


Figure 11. Graphical Representaion of Pairing Size and comutational cost in Pairing strategy

Figure 12 shows the tradeoff between window size and number of pre computations in Pairing strategy.

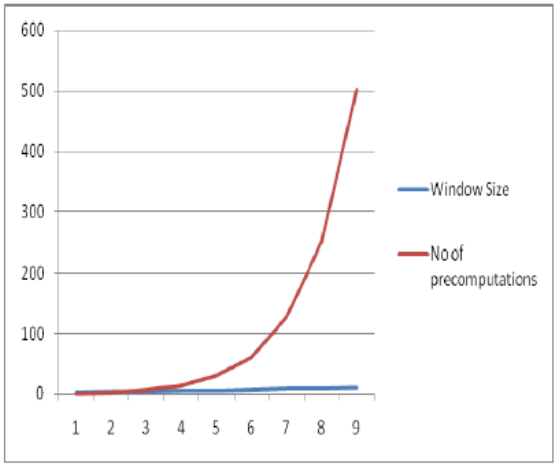


Figure 12. Graphical Representaion of Pairing Size and number of pre computation in Pairing strategy

TABLE III.

PAIRING SIZE VS NO OF DOUBLINGS, ADDITIONS AND PRE COMPUTATIONS IN 1'S COMPLEMENT BASED ARITHMETIC STRATEGY

PSIZE	Number of Doublings	Number of Additions	Number of Pre-computation
2	9	4	1
3	7	3	3
4	6	2	7
5	5	1	15
6	4	1	31
7	3	1	61
8	3	1	127
9	1	1	251
10	0	0	501

TABLE IV. EXPERIMENTAL RESULTS OF SOFT COMPUTING BASED ARITHMETIC STRATEGY (KB- KILO BYTES, MS- MILLE SECONDS, E-ENCRYPTION TIME, D-DECRYPTION TIME, AND T- TOTAL TIME)

File size (KB)	ECC without Soft Computing based Arithmetic (Time in ms)			ECC with without Soft Computing based Arithmetic (Time in ms)		
	E	*D*	*T*	*E*	*D*	*T*
1	1490	1370	2860	1320	1340	2660
3	2470	2410	4880	1920	1980	3900
5	3510	3460	6970	2820	2890	5710
7	4573	4512	9085	3916	3996	7912
8	5685	5642	11325	4613	4663	9276

From table IV, it is observed that the time taken for encryption and decryption with fuzzy (soft computing strategy) is substantially decreased compared to corresponding counterpart without fuzzy.

Figure 13 shows the output with StorageRoom and PreComputing for fuzzy control in ECC scheme.

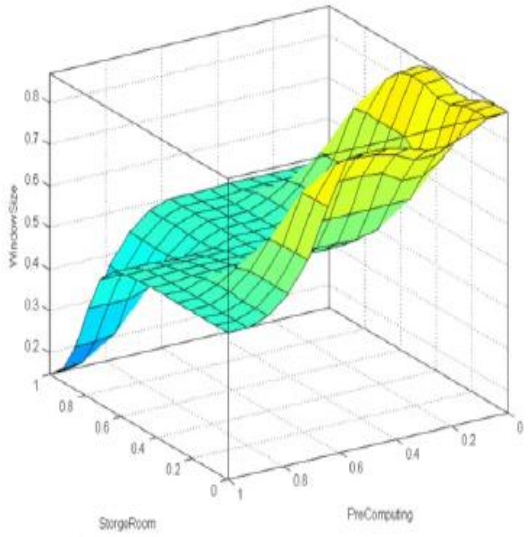


Figure 13. Graphical Representaion of output of the exterior StorageRoom vs. PreComputing for fuzzy control in ECC scheme

From the above figure 13 it is observed that in Min window size side *if* the *(StorageRoom is Min) the conquered function of "doubling" will play role*. But if the window size is at the Max side, the StorageRoom will be moderately stayed at the middle either for PreComputing or Doubling. Doubling will stridently increase when window size is little bit larger.

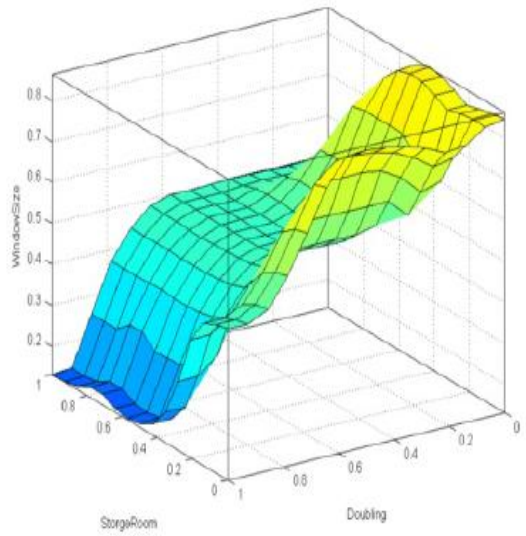


Figure 14. Graphical Representaion of output of the exterior StorageRoom vs. Doubling for fuzzy control in ECC scheme.

In figure 14 it is observed that larger window size for doubling is needed when StorageRoom is Max. Now set the weight value equals to 0.5 for the fuzzy rules 1, 5, 10, 13, 14, 15, 16, 18, 20, 21, 22, 23, 25, 26 then doubling will be increase by larger window size due to the fact that StorageRoom controlled the major functions. If PreComputing and Doubling are decreased by 0.02% and StorageRoom is increased by 0.04% then the output will change.

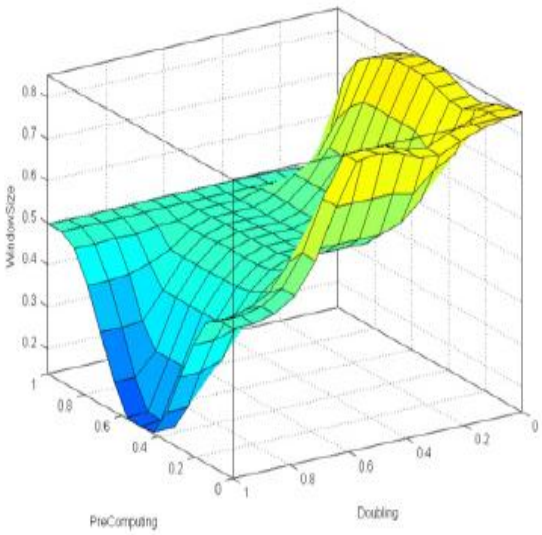


Figure 15. Graphical Representaion of output of the exterior StorageRoom =0.4 and PreComputing vs. Doubling for fuzzy control in ECC scheme

After initializing constant value to the StorageRoom i.e. StorageRoom = 0.4 the output of the PreComputing vs. Doubling is shown in figure 15.

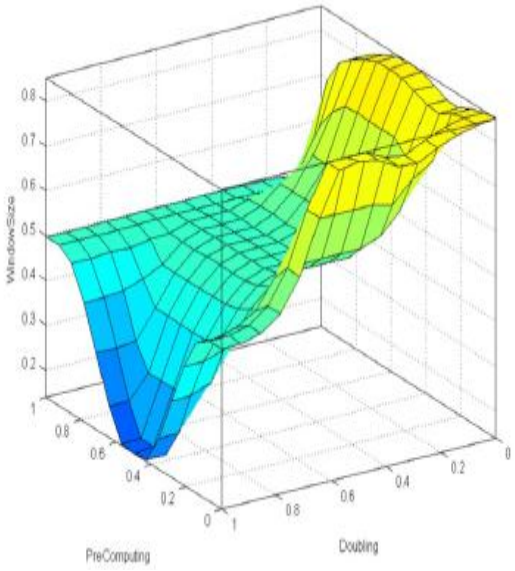


Figure 16. Graphical Representaion of output of the exterior StorageRoom =0.8 and PreComputing vs. Doubling for fuzzy control in ECC scheme

Figure 16 shows output of the PreComputing vs. Doubling with StorageRoom=0.8. This example shows that Storageroom is not dominating factor. The output of the "*PreComputing*" vs. "*Doubling*" is not much difference between Figures 8 and 9. It is noticed from figures 13 and 14 that the "dominated factor" is not the members of the "StorageRoom" in figure 15 due to the fact that the StorageRoom goes to large than 40% (or 0.4) the factors PreComputing and Doubling are dominated in figure 13 and figure 14 separately.

VII. Conclusions And Future Scope

This paper presents a novel FLHSPKC technique which helps to provide a faster version of traditional ECC. In WSN ECC is the best alternative to RSA, DSA and other prime field based cryptographic algorithms. Because ECC offers same security level with fewer amounts of time and memory complexity compared to RSA, DSA. WSN has restraint in the bandwidth, computational potency, power availability or storage. Aim of this paper is to provide some alternative schemes which can be associated with traditional ECC to make it faster and suitable for WSN. Proposed FLHSPKC techniques provide following schemes a) Hamming weight reduction strategy, b) Pairing strategy, c) 1's complement arithmetic strategy, d) Soft Computing based Arithmetic Strategy.

Either of these strategies can be incorporated in traditional ECC for faster arithmetic computation. Soft computing based controller for ECC also been proposed in this paper. Pre-computing is related to the storage i.e. ROM and Computing is associated with the computing capability or capacity which is RAM. In this paper proposed soft computing based controller always controlled the window size which is a trade off between available ROM and RAM in a sensor node for a particular instance of time.

Future scope of this FLHSPKC technique is that

- FLHSPKC technique deals with only integer set. So, special character constant can also be considered with integer set for better security purpose.
- For security reason key length of the proposed method can also be increased.

ACKNOWLEDGMENT

The author expressed deep sense of gratitude to the Department of Science & Technology (DST) , Govt. of India, for financial assistance through INSPIRE Fellowship leading for a PhD work under which this work has been carried out, at the department of Computer Science & Engineering, University of Kalyani.

REFERENCES

[1] W. Diffie, M.E. Hellman, New directions in cryptography, IEEE Transactions on Information Theory 22 (6) (1976) 644–654. 1976.

[2] N. Koblitz, "Elliptic Curve Cryptosystems", Mathematics of Computation Vol. 48 (177), pp. 203-209, November, 1982.

[3] T. ElGamal, Public key cryptosystem and a signature scheme based on discrete logarithms, IEEE Transactions on Information Theory 31 (4) (1985) 469–472, 1985.

[4] N.Koblitz, "Elliptic Curve Cryptosystems," Mathematics of Computation, vol. 48, pp. 203-209, 1987.

[5] K. Koyama and Y. Tsuruoka.1993 Speeding up elliptic curve cryptosystems by using a signed binary window method. Advances in Cryptology – Crypto '92, LNCS, 740, Springer- Verlag, 345-357, 1992.

[6] D. Knuth, "The Art of Computer Programming – Semi Numerical Algorithms", Vol.2, Addison-Wesley, Third Edition, 1998.

[7] N. Koblitz, "A Course in Number Theory and Cryptography", Second Edition, Springer-Verlag, 1994.

[8] J. Kelsey, B. Schneier, D. Wagner, C. Hall (1998). "Cryptanalytic Attacks on Pseudorandom Number Generators". Fast Software Encryption, 5th International Proceedings, 1998.

[9] Cohen H., Miyaji A., and Ono T. 1998. Efficient elliptic curve exponentiation using mixed coordinates. In Proceedings of the International Conference on the Theory and Applications of Cryptology and Information Security (ASIACRYPT'98). Springer-Verlag, London, UK, 51–65, 1998.

[10] J. L´opez and R. Dahab.1999 Improved Algorithms for Elliptic Curve Arithmetic in GF(2n). Selected Areas in Crytography - *SAC '98*, LNCS 1556, Springer-Verlag,201-212, 1998.

[11] S. Shantz,2000. From Euclid's GCD to Montgomery Multiplication to the Great Divide, preprint, 2000.

[12] cryptography and RSA on 8-bit CPUs," in Proceedings of the 2004 Workshop on Cryptographic Hardware and Embedded Systems (CHES) August , 2004.

[13] Malan D. J.,Welsh M., and Smith M. D. A public-key infrastructure for key distribution in TinyOS based on elliptic curve cryptography. In Proceedings of the 1st IEEE International Conference on Sensor and Ad Hoc Communications and Networks. Santa Clara, CA. October, 2004.

[14] N. Gura, A. Patel, A. Wander, H. Eberle and S. C. Shantz,2004.Comparing elliptic curve cryptography and RSA on 8-bit CPUs, In "Proceedings of Workshop on Cryptographic Hardware and Embedded Systems (CHES'04)," Springer, 119– 132, 2004.

[15] Yong-Je Choi, Moo-Seop Kim, Hang-Rok Lee, and Ho-Wan Kim, Implementation and analysis of Elliptic Curve Cryptosystems over Polynomial basis and ONB", PWAEST, Vol. 10, pp. 130-134, December, 2005.

[16] Atul Kahate, Cryptography and Network Security, 2003, Tata McGraw-Hill publishing Company Limited, Eighth reprint 2006.

[17] G. Werner-Allen, K. Lorincz, M. Ruiz, O. Marcillo, J. Johnson, J. Lees, and M. Welsh, "Deploying a wireless sensor network on an active volcano," IEEE Internet Computing, vol. 10, pp. 18-25, 2006.

[18] Sarkar Arindam, Mandal J. K., " Energy Efficient Wireless Communication using Genetic Algorithm Guided Faster Light Weight Digital Signature (GADSA)", International Journal of Advanced Smart Sensor Network Systems (IJASSN), Vol 2, No.3, July 2012, DOI: 10.5121/ijassn.2012.2302, pp. 9-25. ISSN: 2231 - 4482 [Online]; 2231 - 5225 [Print], 2012.

[19] Sarkar Arindam, Mandal J. K., " Multilayer Perceptron Guided Key Generation Through Mutation with Recursive Replacement in Wireless Communication (MLPKG)", International Journal on AdHoc Networking Systems (IJANS) Vol. 2, No. 3, July 2012, DOI: 10.5121/ijans.2012.2302, pp 11-28, ISSN: 2249 - 0175 [Online]; 2249 - 2682 [Print], 2012.

[20] Sarkar Arindam, Mandal J. K., "Swarm Intelligence based Faster Public-Key Cryptography in Wireless Communication (SIFPKC)", International Journal of Computer Science & Engineering Technology (IJCSET), Vol. 3 No. 7 July 2012, pp 267-273, ISSN: 2229-3345, 2012.

[21] Sarkar Arindam, Mandal J. K., "Secured Wireless Communication by High-Speed RSA Using Evolutionary Programming based Optimized Computation (HS-RSA-EP)", International Journal of Advanced Research in Computer Science (IJARCS) 2012.

AUTHORS PROFILE

Arindam Sarkar

INSPIRE Fellow (DST, Govt. of India) at the department of Computer Science & Engineering, University of Kalyani, MCA (VISVA BHARATI, Santiniketan, University First Class First Rank Holder), M.Tech (CSE, K.U, University First Class First Rank Holder). Total number of publications 14.

Jyotsna Kumar Mandal

M. Tech.(Computer Science, University of Calcutta), Ph.D.(Engg., Jadavpur University) in the field of Data Compression and Error Correction Techniques, Professor in Computer Science and Engineering, University of Kalyani, India. Life Member of Computer Society of India since 1992 and life member of cryptology Research Society of India. Dean Faculty of Engineering, Technology & Management, working in the field of Network Security, Steganography, Remote Sensing & GIS Application, Image Processing. 25 years of teaching and research experiences. Eight Scholars awarded Ph.D. one submitted and 8 are pursuing. Total number of publications 252.

A Hindi Speech Actuated Computer Interface for Web Search

Kamlesh Sharma
Research Scholar, Dept. of CSE
Lingaya's University
Faridabad, Haryana, India

Dr. S.V.A.V. Prasad
Dean of R&D
Lingaya`s University
Faridabad, Haryana, India

Dr. T. V. Prasad
Dean of Computing Sciences
Visvodaya Technical Academy
Kavali, Andhra Pradesh, India

***Abstract*— Aiming at increasing system simplicity and flexibility, an audio evoked based system was developed by integrating simplified headphone and user-friendly software design. This paper describes a Hindi Speech Actuated Computer Interface for Web search (HSACIWS), which accepts spoken queries in Hindi language and provides the search result on the screen. This system recognizes spoken queries by large vocabulary continuous speech recognition (LVCSR), retrieves relevant document by text retrieval, and provides the search result on the Web by the integration of the Web and the voice systems. The LVCSR in this system showed enough performance levels for speech with acoustic and language models derived from a query corpus with target contents.**

Keywords- Web search; Hindi speech; HSACIWS; computer interface; human computer interaction.

I. INTRODUCTION

Information contained on the World Wide Web is inaccessible to many people. The web is primarily a visual medium that requires a keyboard and mouse to navigate, and this disenfranchises several types of users. People who lack in skills to use a keyboard and mouse find navigation difficult. Handicapped users also cannot use keyboard and mouse and people who do not have access to an Internet-capable computer have difficulty even accessing the Web. Speech recognition and generation technologies offer a potential solution to these problems by augmenting the capabilities of a web browser.

In the present development of human computer interaction (HCI), Automatic Speech Recognition interface (ASRI) is an emerging technology for offering users a totally new way of mouse, keyboard and joystick control, by using speech [1]. ASRI systems can provide both typing and web browsing for the disabled who cannot use real mouse and keyboard to obtain information from Internet as normal person do. HSACIWS users can input the query in the form of speech on the web page of the browser by using headphone. After confirmation, HSACIWS can capture the uniform resource locator (URL) and open the web page they are interested in. This novel application integrates functions of character input and cursor control, and may help those disabled to obtain information by using the search engines. World Wide Web as a repository of information for unconstrained and wide dissemination, information is now broadly available over the internet and is accessible from remote sites.

II. NEED OF INTERFACE

As network and internet have become popular and used by everyone around the word for accessing the data stored in storage mediums attached to network and servers on the internet. Data could be text, images, video, audio and other representation and non representation information. Search engines and directories were used in making queries and searches of stored data then returning the result of the query to the user. Currently used search engines and directories Google, Yahoo, Alta Vista, Ask, MSN Search, AOL, Lycos, Looks Smart and other search engine provide their search service via servers connected to networks and internet. [2]

The search engine primary mechanism is to navigate to a web page requested by user when the page is stored on any server in internet. The users manually type characters words or phrases known as a query into the search engine form. No search engine currently offers an implementation solution allowing user to make queries by speaking the query term in Hindi voice, converting the Hindi speech words into data then converting them to English word query and finally processing the query to perform a search.[3]

Speech recognition technology was developed over the past three decades. It is used in many fields like automatic speech recognition directory, military, defense, medical science, bio-informatics, home automation systems, word processing system, dictation system, embedded systems, query processing and many more systems developed for handicapped persons [7]. Most of the systems work on English language and they function at variable level of effectiveness due to limitation of the software understanding and the complexity and the variability of human speech. Large numbers of methods have been developed to acquire the accuracy in English language.

In order to enable a wider proportion of the population of India to benefit from the internet, there is a need to provide additional interface between the user and machine. Speech being a natural mode of communication among human beings can also be a convenient mode of interaction with a computer. Internationally, efforts are already on to combine hypertext navigation with spoken language. This is of particular significance in India where the rate and level of literacy are quite low. Coupled innovatively with visuals, speech and sound can add a new dimension for conveying information to the common man. It is desirable that the human-machine

interface permits communication in one's native language. This is an important issue in a multi-lingual country such as India where large numbers of people are comfortable with Hindi language. If the human oriented information over the internet can be access in Hindi language, the computer can process such hypermedia document and provide the information appropriately to a large number of users [14].

In India, Hindi speech recognition is upcoming field of research and people are working in many aspect of Hindi speech recognition. The Hindi speech engine could be utilized in search engine for querying will make the learning and use very easy for the common man. Such type of interfaces can be used by handicapped also and most importantly the person need not be aware about how to use computers.

III. System Design

The manner in which users interact with a program is known as its user interface. The user interface controls how data is entered and how information is displayed. There are mainly five types of user interfaces: character user interface (CUI), graphical user interface (GUI), touch user interface (TUI), voice user interface (VUI) and brain signal user interface (BSUI). Herein a voice user interface comes into existence. Fig. 2 shows a flowchart of Hindi speech actuated computer interface for web search shows the system design of speech actuated interface.

As shown in Fig. 1, the speech actuated computer interface consists of wearable headset and a laptop computer or desktop computer. Headphone is used to take the speech query from user. A good quality headphone acquired, amplified and digitalized the Audio signal, then wirelessly or transmitted it to the computer. The software running on laptop computer analyzed the signal and performed the instructions issued by the user. The voices utterances are of words fed to statistical speech recognition model using Hidden Markov model (HMM) where the word that were utters most likely are determined. A database was constructed with a list of word defining specific subject like fruits, vegetable, news, recipes, stock, weather etc. [9]

The uttered words are compared to the database words, if uttered word match is found a set of keywords are formed to make a query. This query is input to a search engine and the search engine processing the query and returns the result.

A user would use a speech actuated user interface in which user query is received through a device in signal form of speech and converts these speech signals to digital signals. The digital signals received as input to speech recognition module accepts the natural continuous speech patterns and generates most probable words uttered by using the HMM. The output from speech recognition module is searched against a large database of words stored in previously formulated and trained database. The output from speech recognition module is Hindi words send as input to "Speech conversion from Hindi to English" module which converts these Hindi words into corresponding English words.

The strings of English words is passed to a "Find out keywords and drop the stop constant" module for processing as in Fig. 2. Keywords are identified and marked and that recognized speech which was not useful or will not be used for search query were dropped. The keywords passed to "Make a Query" module formed a new query with the keyword and passed to "Search engine module". In case keywords are not found then the query was passed again through the user Hindi query module. The search engine searched the pages on the internet and fetched the pages on the client machine and finally displayed the results on user's screen. In case results were not found then it displayed a message on user's screen suitably.

The forming of user query was basically a combination of words that were identified from utterance using HMM. HMM [5] was used for representing speech units. Forty eight phone-like acoustic-phonetic units were used to represent Hindi sentences. Only monophone models were used due to paucity of sufficient speech data. The plosives were represented as two units-the first representing the closure, and the second representing the rest of the stop consonant. While distinction was maintained between the releases of aspirated and unaspirated plosives, no such distinction was retained between their closures. In addition to the Hindi phonemes, a few commonly occurring English vowels such as /ae/ were included. At the model level, gender identification technique can be used and the test feature vector sequence can be matched with a gender dependent HMM. Initialisation of emission probability densities of HMM states using a segmented and labelled speech database [6] should lead to better models.

Fig. 1 User interacting with system without mouse and keyboard

Dropping of stop constant and identification of key words was very important to the searching process, as shown in Fig. 2. Search engine returned results on the basis of query formed with the help of keywords identified by find keywords module.

For instance, if user gave a query for processing to the HSACIWS is "*Aaj Dili ki mandi mein aalu ka bhav kya hai*". This query was given as user query in Hindi module. The recognized string of words send as input to speech conversion from Hindi to English module which converts the Hindi string to corresponding string with the help word model and morphological analyzer is "what is the price of potatoes in the market of Delhi today". This converted string is passed as input to find the keywords and dropped the stop constant

module which tokenized the keywords like "today", "what", "is", "the", "price", "of", "mango", "in", "Delhi", etc. and drop the stop constants. These keywords were passed as input to make a "Make a query" module which applied preprocessing techniques to generate a suitable query for search engine.

This was for HSACIWS to make sure that the input query was right as recognition was error prone. The system needed to identify the errors and came up with appropriate strategies to overcome the errors. A query was identifying in different ways as shown in Table 1. The given query "*Aaj Dili ki mandi mein Aalu ka Bhav kya hai*" could be recognized as:

Table 1. Query recognition with different variation in keywords

Aaj Bili ki mandi mein Aalu ka Bhav kya hai
Aaj Dili ki dandi mein Aalu ka Bhav kya hai
Aaj Dili ki mandi mein Balu ka Bhav kya hai
Aaj Dili ki mandi mein Balu ka Bhav kya hai
Aaj Bili ki mandi mein Balu ka Bhav kya hai
Aaj Dili ki mandi mein Aalu ka kya hai
Aaj Dili ki mandi mein ka Bhav kya hai

In such type of cases when keywords were not recognized correctly, control was passed back to user query in Hindi module and the user repeated the query again.

The results were displayed but in many cases was not sufficient to answer the user. At this point the user was allowed to optionally select to run a different module to mark hyper text links or filter information from hyper links related to the user query. Depending on the user's choice, the results could then be displayed as text on the screen or played as speech. The output can be seen on screen in two ways, viz., (a) hyperlink number technique as shown in Fig. 3 or (b) employing a data filtering algorithm to get the actual results [10].

In hyperlink number technique a visual numbering was provided to the hyperlinks and indicating to the user how to activate them. The user can activate the hyperlink number module by providing visual number to the module [7]. Data is filtered and processed using a data filtering algorithm according to the user queries and results are sent to the speech synthesizer which translates the result from text to speech before playing it on the client system. [4]

The following are the major issues involved in the development of HSACIWS:

- *Noisy environment:* The target users of HSACIWS were primarily the farmers in rural and semi urban areas. The quality of speech signal was affected by the distance of microphone, environment where system was placed, speech codec's and communication channel. HSACIWS system was expected to work in noisy environments, including background speech.
- *Dialect/Pronunciation variation:* The user spoke in different styles. Each dialect differed from the other at phonetic, phonological, morphological, grammatical and lexical levels. Some time moods (anger, illness, happiness and sadness) of the users effected the dialect/pronunciation variation.
- *Unstructured conversation:* The target audience of the HSACIWS may not have interacted with a computer based information access system. Hence, the conversation was typically unstructured and was filled with inconsistencies including repeats and false starts.

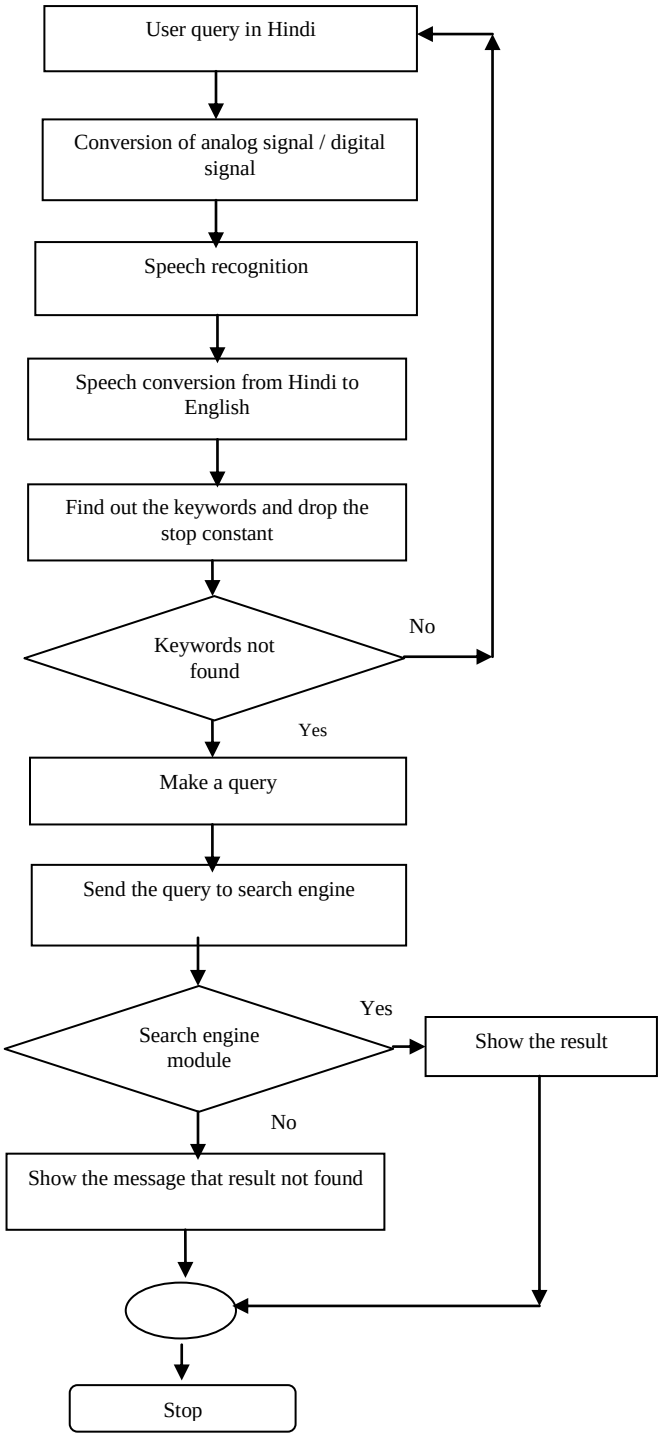


Fig. 2 Flowchart of Hindi speech actuated computer interface for web search

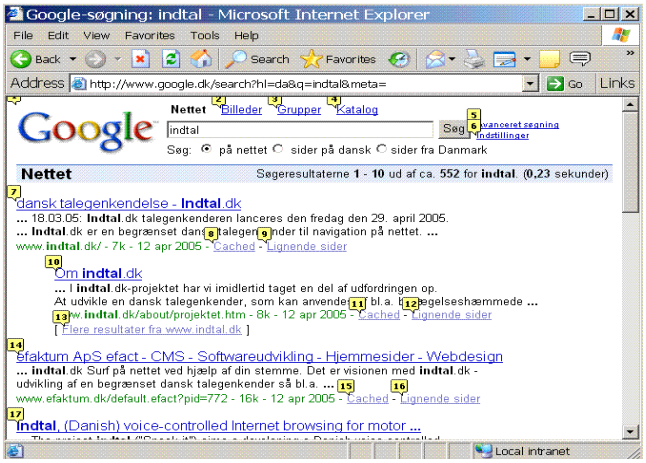


Fig. 3 Displaying hyper links with number in the web Browser [8]

A. *Training Module*

The system was trained by actual speakers articulating word continuously. Continuous speech was marked by sounds or phonemes that were connected to each other. The audio signal was processed and features extracted. The signal was smoothed by using different filters to form feature vector which were computed periodically, say every 10 to 20 milliseconds. Many types of features were used including time and frequency masking, tasking of inverse Fourier transforms resulting in a mathematical series of which the coefficient were retained as feature vector. The features were handled mathematically as vectors to simplify the training and recognition computation.

Fig. 4 describes an approach to trained the system for reading the spoken words used in the training were listed in a lexicon and a phonetic word model by using the HMM were from lexicon and phonetic spelling. These HMM word models were iteratively compared to the training speech. Training speech was produced by these HMM words models and the grammar was established with the lexicon a single probabilistic grammar for the sequence of phonemes was formed and stored in dictionary.

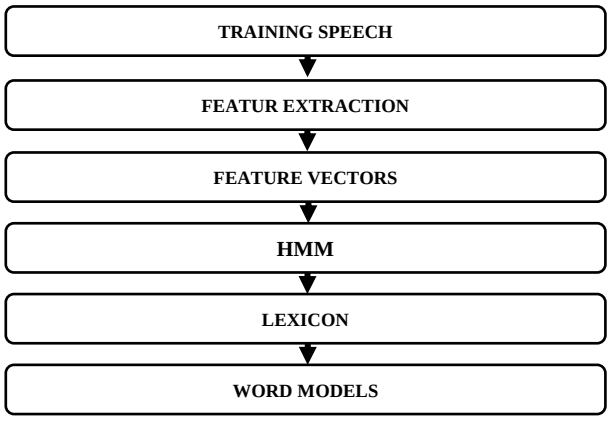


Fig. 4. Functional block diagram of the Training Module

B. *Speech Recognition Module*

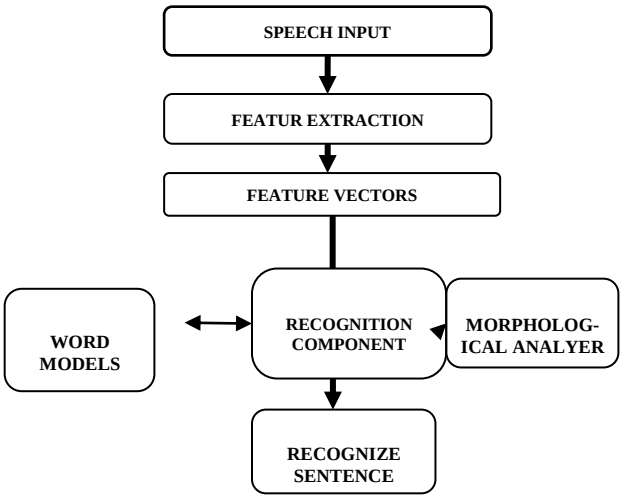


Fig. 5. Functional block diagram of the Speech Recognition Module

Recognition of an unknown speech begins with extracting feature and generating a feature vector for a particular speech by using different vectors generating technique as discussed in training module. Output of feature vector was passed to recognition component as shown in Fig. 5. Recognition component used HMM model sequences allowed by the grammar were searched to find the word sequence in the word model with the highest probability of generating that particular sequence of feature vector.

C. *Morphological Analyzer*

Morphological analyer component was used to generate a complete sentence with the help of words recognized by recognized component using the reverse and forward methodology. For translation from one language to other, the source language was first analyzed for finding the required attributes. In the source sentence the words may exist in any of their forms, so we first found their root words and then other attributes. Finding root words in the source language is called Reverse Morphology, also known as Morphological Analyzer. For the target language, the words from the given root word and their attributes were generated, and hence called forward morphology. The Morphological Analyzer constitutes of following sub modules: Input Module, Input Normalizer Module, and Tagger Module, as shown in Fig. 6 and which are described below: [13]

a) *Input Normalizer:* Input Normalizer separates the entered text into the words. It separates the words and stores them whenever blank space is encountered and provides it as input to other modules. This module also searches for the presence of the auxiliary words in the sentence and removes them, if present.

The auxiliary and its attributes are stored for further processing. The presence of auxiliary is language dependent as some languages like Sanskrit does not have auxiliaries. This module only removes the extra spaces between the words and sends the normalized sentence for further processing. [11]

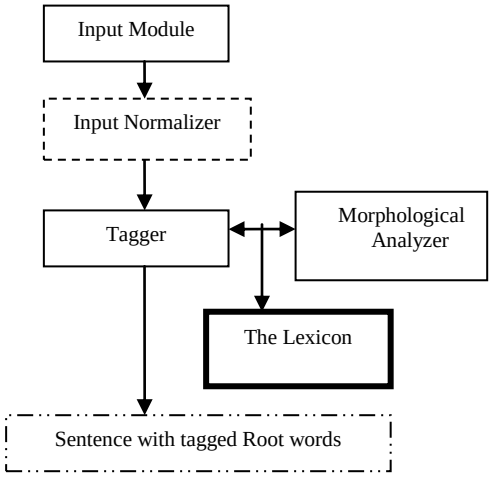


Fig. 6. Functional block diagram of the Morphological Analyzer

b) Tagger: The tagger takes the normalized sentence from the previous module and subdivided into it lexical items. Lexical items are not necessarily single words. More than one word in the input sentence may form a lexical item (e.g. give up, put off, etc). The process of dividing the sentence into lexical items is often known as Lexical Analysis. Given "What is the price of potato in Delhi market today" as input, this module would tokenize the words into an array of lexical items. Once the lexical items are obtained the next task of the tagger is to obtain the category and subcategory information for each of these items. It uses morphological analyzer and source lexicon to obtain the category and subcategory information for the lexical items.

c) Source Language Lexicon: The bilingual lexicon took Hindi as the source language and English as the target language. The lexicon contained the categorization and sub categorization information about source and target words, as in Table 2. For example:

Table 2. Categorization information

Category	Subcategory information stored in lexicon
Noun	gender, number, person, noun case
Verb	tense, aspect, gender

A given word may carry different categories and subcategories information both in the source as well as in the target language. Further, there may not be exact matching of attributes from the source language to the target language. For instance, the verb in Hindi language has two numbers (singular, plural) and two genders (masculine, feminine) whereas English language verbs does not have gender. The lexicon of a language is its vocabulary, including its words and expressions. More formally, it is a language's inventory of lexemes.

The lexicon includes the lexemes used to actualize words. Lexemes are formed according to morpho-syntactic rules and express sememes. In this sense, a lexicon organizes the mental vocabulary in a speaker's mind: First, it organizes the vocabulary of a language according to certain principles (for instance, all verbs of motion may be linked in a lexical network) and second, it contains a generative device producing (new) simple and complex words according to certain lexical rules. There were five tables a) Hindi root lexicon, b) Hindi verb feature, c) morphological lexicon, d) noun feature and e) suffices. Usually a lexicon is a container for words belonging to the same language. [12]

d) Morphological Analyzer: This module is a part of tagger which finds out category and sub-categorical information of the lexical items. The Morphological Analyzer is an integral part of any Natural Language Processing system, especially in the context of Indian languages. For fixed word order languages, the semantics of a word are primarily governed by its absolute and relative position inside a sentence. However, for free word-order languages, any clues about the semantics cannot be obtained from its position in the sentence. In case of Indian languages, which are mostly free order (like Hindi); the semantics (part of speech and other subtleties) are heavily dependent on the surface structure of the word. The task of the morphological analyzer is to identify the structural components of a word, and hence glean information about it. [13]

When an input sentence was fed for translation, the morphological analyzer identified the proper words in the sentence and retrieved necessary information about those words from the lexical database. Lexical database stored only the root form of words and its syntactic and semantic information. With the help of paradigm files, root word was extracted from the original word and all the information about that word was retried.

IV. Results And Discussions

A. Training and Testing Data Scenario for Experiments

In order to compare the effectiveness of the HSACIWS system under a scenario of truly limited data resources. There is a collection of training and testing data for the HSACIWS system that contained extremely limited amounts of data. The Training and testing data was extracted from the Indian urban and semi urban areas that have been collected through the questionnaires, Templates and personal talk. The large number of data has been collected through the questionnaires where the users were asked for different question for which the system could train and test.

The limited data consisted of the following resources:

a. *Data Corpus:* The database contains 478 word-aligned phrases and sentences from the user in urban and semi

urban areas. This elicited data collection includes both the training and testing phrases and sentences.

b. *Small Hindi-to-English Lexicon:* The database contained 2390 Hindi to English translation pairs in database. The "Speech conversion from Hindi to English" module of the system used the database and runs each Hindi input word through the morphological analyzer. The Morpher returns the root form of the word, along with a set of morphological features. The root form is then matched against the lexicon, and the set of lexical transfer rules that match the Hindi root are extracted. The morphological features of the input word are then unified with the feature constraints that appear in each of the candidate lexical transfer rules, pruning out the rules that have inconsistent features and generate the English sentence.

c. *Small English-to-Hindi Lexicon:* The database contained 2105 English to Hindi pairs in database. The "Show result" module uses the database to generate the result from English to Hindi.

B. Experimental Testing

Experiments were conducted to evaluate the baseline system and the improved HSACIWS. The HMM Toolkit was used in this experiments. A set of 100 sentences were randomly chosen from the set of 478 training sentences for test the HSACIWS. A set of 20 subjects were randomly chosen from 100 subjects (students, farmers, housewives and teachers) and were asked to raise a query to the HSACIWS to retrieve the result accordingly and show the user. The user were not trained or provided any information. Initially the experiment is done on pre define queries. The performance of the HSACIWS was calculated by using the percentage performance formula Accuracy which is defined as

$$\text{Accuracy} = \frac{S - E_s - E_d}{S} \times 100\%$$

where S, E_s, and E_d denotes the total number of sentences in the test sentences, the number of substitution errors, and deletion error respectively. The Accuracy as 100 times the ratio of the number of complete sentences recognized correctly to the total number of sentences in the test suite and show the result.

The success of each trial was based on whether the system was able to retrieve the required information to the user or not. For example user will raises the queries to the HSACIWS and system respond accordingly.

User: Sone ka Bhav kya hain.

System: Sone ka Bhav 31500 rupai hai.

User: WHO ka kya matlab hai.

System: WHO ka matlab World Health Organization.

User: Bharat ki Rajdhani kya hai.

System: Bharat ki Rajdhani delhi hai.

Over 200 experiments were conducted on the HSACIWS with 20 users (10 male and 10 female) on 100 different queries. The results are as shown in Table 3 and Table 4:

Table 3. Performance analysis of HSACIWS

Data Type	Number of Sentences	User	Sex	Total	Accuracy
Training	478	12	M/F	5736	79.3%
Test	100	10	M	1000	78.9%
Test	100	10	F	1000	77.8%

Table 4. Overall system performance analysis

System	Accuracy
HSACIWS	78.5%

The success of system was based on the confidence whether the system was able to retrieve the required information correctly to the user or not.

V. CONCLUSION

We present initial efforts for utilizing spoken Hindi language as a means of communicating web information to common Indians. We presented a voice query retrieval system in Hindi applied to document search on Internet or network.

The results of this experiment suggest that native Indians who are not able to use the computer and/or lack English skills will be able to use voice based control to navigate and obtained responses from the World Wide Web. They will not need to train the speech recognition software to their specific voice.

This system provides access to digital content over the internet to illiterate, vision-impaired, urban and semi-urban Indian people who are not able to read/write English language.

The same Speech interface can be enhanced to work for different regional languages like Punjabi, Marathi, and Telugu, etc. to enhance or to extend it for different regional languages what all is needed is the transfer or translation rules of grammar, which can be generated with great ease by using the same dataset with different target languages. Hence, there is a need to design and develop special user interfaces for accessing web information by speech of different languages.

REFERENCES

[1] Takahiro Ikeda, Shin-ya Ishikawa, Kiyokazu Miki, Fumihiro Adachi, Ryosuke Isotani, Kenji Satoh and Akitoshi Okumura, Speech-Activated Text Retrieval System for Cellular Phones with Web Browsing Capability, Proceedings of PACLIC 19, the 19th Asia-Pacific Conference on Language, Information and Computation.

[2] Michael D. Goller and Stuart E Goller, Speech Interface for search Engine".United state patent, Jun. 22 2010, shett no 1 to 4.

[3] Frederick J. Damerau and David E. Johnson, Automated set up of web-based Naturel Language Interface, U S Patent, Jun. 22 2010,

[4] M.K. Brown, Grammar Representation Requirements for voice markup Langauge, Bell Labs, Murry Hill, NJ, Dec. 1999.

[5] Honglai Xu, Tianyi Qian, Bo Hong, Xiaorong Gao, and Shangkai Gao, A Brain Actuated Human Computer Interface for Google Search, IEEE 2009.

[6] Matrin Holley and Dieter Kubesch, Speech Recognition Method for Activating a hyperlink of an Internet page, US patent 3 Aug, 2003.

[7] Bruce Moulton, Gauri Pradhan and Zenon Chaczko, Voice Operated Guidance Systems for Vision Impaired People: Investigating a User-Centered Open Source Model in International Journal of Digital Content Technology and its Applications, Volume 3, Number 4, December 2009.

[8] Edward V Porter, Voice Recognition system, U S Patent, 9 May 1989.

[9] Cameron, Hugh: " Speech at the Interface". Proceedings of the Cost249 Workshop on Speech in Telephone Networks, Ghent 2000.

[10] Hadjadj, Djamel and Dominique Burger, "BrailleSurf: An HTML Browser for visually handicapped people". In Proc. of 14th conference on "Technology and Persons with Disabilities", Los Angeles, 2000.

[11] Mihelic, France, Nikola Pavesic, Simon Dobrisek, Jerneja Gros, Bostjan Vesnicer and Janez Zibert: Homer – A Small Self Voicing Web Browser for Blind People Laboratory of Artificial Perception, Systems and Cybernetics Faculty of Electrical Engineering, University of Ljubljana, Slovenia, 2002

[12] Robin, Michael B. and Charles T. Hemphill: Considerations in Producing a Commercial Voice Browser, W3C WS on "Voice Browsers". Massachusetts, 1998.

[13] Verb Morphology for English, available at http://www.xrce.xerox.com/competencies/contentanalysis/demos/doc/mor-eng-2.html

[14] Prahallad Kishore and Black Alan, A text to speech interface for Universal Digital Library, J Zhejiang Univ Science 6A(11):1229-123.Journal of Zhejiang University, 2005

AUTHORS' PROFILE

Ms. Kamlesh Sharmsa received her masters in Computer Sc. & Engg. degree from Maharshi Dayanand University, Rohtak, India in 2009. She is currently associated with at Lingaya's University, Faridabad in the Dept. of Comp. Sc. & Engg. as Research Scholar. She has over 7 years of teaching experience at under graduate and graduate levels. Her areas of interest are artificial intelligence, operating systems, web mining, Database Management Systems, etc.

Dr. S. V. A. V. Prasad has over 30 years of experience in industry and academics. He has received his master's degree in Electronics & Communications Engg. from Andhra University, AP, India. He earned PhD from Andhra University, Waltair, Visakhapatnam, India. He was with leading research and manufacturing companies in New Delhi, India. He also taught at leading institutions like the Delhi College of Engg. (now Delhi Technological University), Delhi for many years.. He has worked as Head of the Department of Electronics & Communications Engg., Dean of Academic Affairs and as Dean of R&D and Industrial Consultancy at Lingaya's University, Faridabad. He has lectured at various forums on subjects related to electronics, communications, audio engineering, signal processing, etc. Prof. Prasad is a member of IEEE, ISTE, etc. His research interests include audio engineering, signal processing, etc.. He has large number of papers in different journals and conferences.

Dr. T. V. Prasad has over 17 years of experience in industry and academics. He received his graduate and master's degree in Computer Science from Nagarjuna University, AP, India. He was with the Bureau of Indian Standards, New Delhi for 11 years as Scientist/Deputy Director. He earned PhD from Jamia Millia Islamia University, New Delhi in the area of computer sciences/ bioinformatics. He has worked as Head of the Department of Computer Science & Engineering, Dean of R&D and Industrial Consultancy and then as Dean of Academic Affairs at Lingaya's University, Faridabad. He is with Visvodaya Technical Academy, Kavali as Dean of Computing Sciences. He has lectured at various international and national forums on subjects related to computing. Prof. Prasad is a member of IEEE, IAENG, Computer Society of India (CSI), Indian Society of Remote Sensing (ISRS) and APBioNet. His research interests include bioinformatics, artificial intelligence (natural language processing, swarm intelligence, robotics, BCI, knowledge representation and retrieval). He has over 75 papers in different journals and conferences, and also has six books and two chapters to his credit.

A Harmony Search Based Algorithm for Detecting Distributed Predicates

Eslam Al Maghayreh
Computer Science Department
Faculty of Information Technology and Computer Science
Yarmouk University, Irbid 21163, Jordan

***Abstract*— Detection of distributed predicates (also referred to as runtime verification) can be used to verify that a particular run of a given distributed program satisfies certain properties (represented as predicates). Consequently, distributed predicates detection techniques can be used to effectively improve the dependability of a given distributed application. Due to concurrency, the detection of distributed predicates can incur significant overhead. Most of the effective techniques developed to solve this problem work efficiently for certain classes of predicates, like conjunctive predicates. In this paper, we have presented a technique based on harmony search to efficiently detect the satisfaction of a predicate under the possibly modality. We have implemented the proposed technique and we have conducted several experiments to demonstrate its effectiveness.**

Keywords- Distributed Systems; Detection of Distributed Predicates; Runtime Verification; Harmony Search; Testing; Debugging.

I. INTRODUCTION

The design and construction of dependable distributed applications is not an easy task. Several techniques have been used in the literature to improve the dependability of distributed applications. Detection of distributed predicates (runtime verification) is one of the techniques that have attracted a great deal of attention in this regard [1], [2], [3], [4], [5], [6].

Runtime verification techniques can be used to verify whether a given run of a distributed application satisfies certain properties or not. Figure 1 depicts the runtime verification environment [7].

Runtime verification can be used to verify a particular implementation rather than verifying a model of the application as is done in model checking. Moreover, in runtime verification, the properties to be verified can be formally specified, i.e. using temporal logic. Consequently, runtime verification is considered more powerful than traditional testing in this regard [1], [7].

In addition to runtime verification, detecting distributed predicates has several applications. The followings are some of these applications:

a) Detection of when a distributed computation enters a stable state. This involves the detection of a condition which once becomes true, remains true indefinitely. For example, termination detection and deadlock detection.

b) Testing and debugging of distributed programs. Any condition that must be true in a correct run of a distributed application can be specified and then its occurrence can be verified. For example, when debugging a distributed mutual exclusion algorithm, it is useful to monitor the system to detect concurrent accesses to the shared resources. Another example is detecting the presence of multiple leaders in distributed leader election.

c) Identifying bottlenecks. For example, detecting posi- tions during a run of a distributed application where more than n processes from some set are simultane- ously blocked.

Concurrency in distributed applications makes the detection of distributed predicates a very hard and expensive task. Consequently, several techniques have been introduced in the literature to reduce the cost of detecting distributed predicates [7], [1], [8]. A brief review of these techniques will be presented in Section III. However, most of these techniques work well for only certain classes of predicates. It has been proved that the problem of verifying whether a run of a distributed program satisfies certain predicate or not is, in general, an NP-complete problem [1].

In this paper, we exploit harmony search in developing a more powerful technique to detect distributed predicates. This technique can work efficiently for predicates under the possibly modality (a predicate under the possibly modality is evaluated to true if it is true in at least one global state [1]). Harmony search is a popular evolutionary algorithm that can be used effectively to solve problems with exponential size search space (as it is the case in distributed predicates' detection).

The remainder of this paper is organized as follows. In section II, we present a formal model of a run of a distributed program. We discuss other related works in section III. A brief introduction to harmony search is provided in section IV. We present the proposed algorithm in section V. Section VI presents the experimental results. Finally, we conclude our work in section VII.

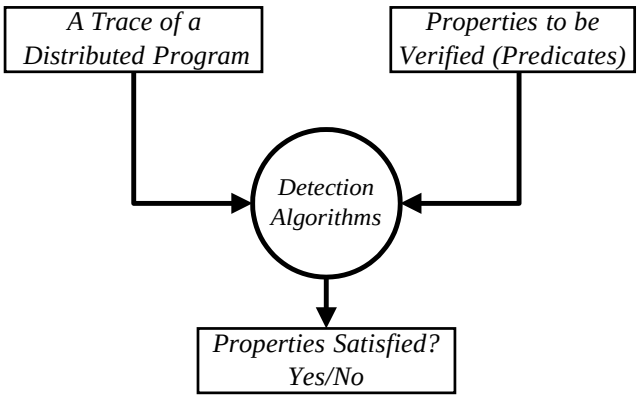


Figure1. Runtime verification environment.

II. Model And Problem Definition

A distributed program consists of n processes denoted by P_0 , P_1 , ..., P_{n-1} and a set of unidirectional channels. An event is the result of executing a statement in a distributed program. An event can be computational event or message event (send/receive). Events are related by their execution order in a process and/or message send/receive relations across processes. The happened-before relation (→) defined by Lamport in [9] applies to all events executed.

***Definition* 1**: A **run** of a distributed program is an event structure <E, →>, where E is the set of events executed, and (→) is the happened-before relation among the events in E.

The space-time diagram shown in Figure 2 depicts a run of a distributed program involving three processes. Time is represented in the horizontal direction and space in the vertical direction. Events are shown as circles in the space-time diagram. Event e_{ij} is the j^{th} event of process Pi. A directed edge is used to link every send event with the corresponding receive event.

The happened-before relation → is a partial order relation on the set of events of any run of a distributed application. $e_{ij} \rightarrow e_{kl}$ if and only if there is a directed path in the corresponding space-time diagram from event e_{ij} to event e_{kl}. If two events e_{ij} and e_{kl} are not related by the happened-before relation, we say that they are **concurrent events** (denoted by $e_{ij} \parallel e_{kl}$). For example, in Figure 2, $e_{01} \parallel e_{21}$ because $\neg(e_{01} \rightarrow e_{21})$ and $\neg(e_{21} \rightarrow e_{01})$.

A consistent cut C of a run $\langle E, \rightarrow \rangle$ is a finite subset of E ($C \subseteq E$) such that if $e_{ij} \in C$ and $e_{kl} \rightarrow e_{ij}$ then $e_{kl} \in C$. The dotted line shown in Figure 2 represents a consistent cut $C1 = \{ e_{01} , e_{11}\}$. Every consistent cut corresponds to a global state of the distributed application represented by the values of the program variables and channels states attained upon the completion of the execution of the events in that consistent cut. The set of global states of a given run endowed with set union and set intersection operations forms a distributive lattice, referred to as the state lattice [10].

Figure 3 depicts the state lattice associated with the run depicted in Figure 2. Each global state can be labeled by the most recent event executed in each process upon reaching it. For example, $(e_{01} , e_{13} , e_{22})$ is the state reached after executing event e_{01} in P_0 , event e_{13} in P_1 and event e_{22} in P_2 . Using the state lattice, we can verify whether a run of a given distributed program satisfies the necessary properties or not (Distributed Predicates Detection).

A predicate is called distributed predicate if the variables involved in expressing the predicate belongs to more than one process. Consequently, the evaluation of a distributed predicate requires the collection of information from several processes. A predicate that involves variables of a single process is called local predicate.

possibly: φ is true if the predicate φ is evaluate to true in at least one global state in the state lattice. definitely: φ is true if, for all paths from the initial global state to the final global state, φ is true in at least one global state along that path [11], [12]. In this paper, we will consider predicates under the possibly modality.

The difficulty of the detection of a distributed predicate is due to the following characteristics of a distributed application [1]:

1) There is no global clock in a distributed application. Consequently, the events of a given run of a distributed application can only be partially ordered.

2) The processes of a distributed application do not have shared memory. As a result, collecting the information necessary to detect a predicate will incur significant message overhead.

3) In any distributed application, there will be several processes running concurrently. As a result, the number of global tates that must be considered to detect a predicate will be exponential in number of processes.

In fact it has been proved that it is, in general, NP-complete to detect a distributed predicate in a run of a distributed application [1]. The next section is dedicated to explore the techniques presented in the literature to detect a predicate in a run of a given distributed application.

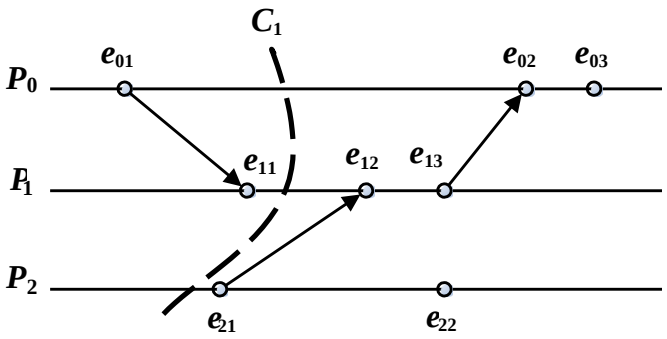


Figure2. A space-time diagram corresponding to a run of a distributed program.

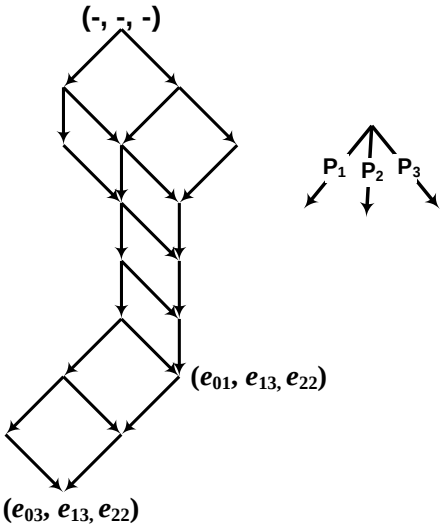


Figure 3. The state lattice corresponding to the run shown in Figure 2.

III. Related Works

Three main approaches have been presented in the literature to detect a predicate in a run of a distributed application. The first approach exploits the global snapshot algorithm proposed by Chandy and Lamport [13], [14], [15]. In this approach, a global state of a run is captured and the predicate to be detected is evaluated on it, if the predicate is evaluated to false, another global state will be captured. This process will continue until we found a global state satisfying the desired predicate. This approach is applicable for stable predicates (once they become true they will not turn false).

The second distributed predicates detection approach was proposed by Cooper and Marzullo [11]. This approach can be used to detect both stable and unstable predicates. It can be used to detect possibly: φ and definitely: φ. However, the detection is very expensive because it requires the construction of the entire state lattice and exploring m^n global states in the worst case, where n is the number of processes and m is the number of local states in each process.

To avoid the construction of the state lattice, the third detection approach exploits the structure of the predicate to identify a subset of the global states, such that if the predicate is true, it must be true in one of the states in this subset. This approach is not as general as the second approach, but it can be used to develop more efficient algorithms for certain classes of predicates. Garg and Waldecker [5], [6] have proposed algorithms of complexity $O(n^2 m)$ to detect possibly : φ and definitely : φ when φ is a conjunction of local predicates.

Several techniques have been introduced in the literature to reduce the cost of detecting a predicate. Computation slicing is one of these techniques. It was introduced in [16], [17], [18], [19] as an abstraction technique for analyzing distributed computations (finite execution traces). A computation slice, defined with respect to a distributed predicate, is the computation with the minimum number of global states that contains all global states satisfying the predicate.

Computation slicing can be used to eliminate the irrelevant global states of the original computation, and keep only the states that are relevant for our purpose. In [19], Mittal and Garg proved that a slice exists for all global predicates. However, it is, in general, NP-complete to compute the slice. They developed efficient algorithms to compute slices for special classes of predicates.

Lamport has presented a theorem on atomicity to simplify verification of distributed and parallel systems [20]. According to this theorem, a sequence of statements in a distributed program can be grouped together to form an atomic action under some stated conditions. An atomic action can receive information from other processes, followed by at most

one externally visible event (for example, changing the value of a variable involved in one of the properties to be verified) before sending information to other processes. Based on this theorem, a distributed program can be abstracted and hence the cost of verifying it can also be reduced.

In [21], [22], [23], the authors have formally defined the notion of atomic actions in message-passing distributed programs. They have exploited the atomicity concept in reducing the state space to be considered in runtime ver- ification of message-passing distributed programs.

In this paper, we exploit harmony search in developing a general and at the same time efficient detection algorithm of distributed predicates under the possibly modality. In the following two sections we will present a brief introduction to harmony search and we will present the details of our proposed detection algorithm.

IV. Introduction To Harmony Search

Harmony Search (HS) is metaheuristic algorithm (also known as an evolutionary algorithm) that emulates the improvisation behavior of musicians [24]. In the music improvisation process, each musician (decision variable) plays (generates) a note (value) for finding a best harmony (global optimum). HS can handle discrete variables [25] as well as contin- uous variables [26], [27]. It has been successfully applied to a wide variety of both discrete and continuous practical optimization problems such as game scheduling problems [28], [29], clustering problems [30], timetabling problems [31], and structural design [32].

The HS algorithm consists of the following main steps [24]:

Step 1. Initialization of the optimization problem and algo- rithm parameters:

In this step, the optimization problem is characterized as a function $\mathbf{f}$ to be optimized (minimize or maximize). Additionally, the control parameters of the Harmony Search are specified in this step including: the Harmony Memory Size (HMS); Harmony Memory Consideration Rate (HMCR);

Pitch Adjusting Rate (PAR); and the stop criterion (i.e. Number of Improvisations (Iterations)).

Step 2. Harmony Memory (HM) initialization:

As shown in (1), the HM composed of HMS (Harmony Memory Size) candidate solutions with N decision variables

$\mathbf{x}_i = [x_i^1, \ldots, x_i^N], \; i \in \{1, \ldots, \mathrm{HMS}\}$. The objective function $\boldsymbol{f}$ in (1) measures the solution quality.

$$HM = \begin{pmatrix} x_1^1 & x_1^2 & \ldots & x_1^N & | & f(\mathbf{x}_1) \\ x_2^1 & x_2^2 & \ldots & x_2^N & | & f(\mathbf{x}_2) \\ \vdots & \vdots & \ldots & \vdots & & \vdots \\ x_{HMS}^1 & x_{HMS}^2 & \ldots & x_{HMS}^N & & |f(\mathbf{x}_{HMS}) \end{pmatrix} \quad (1)$$

In this step, the HM is randomly initialized within the solution space.

Step 3. New Harmony improvisation:

In this step, a new harmony vector $\mathbf{x}_{new} = (x^1_{new}, x^2_{new}, \ldots, x^N_{new})$ is generated based on three operators: (1) memory consideration, (2) pitch adjustment, and (3) random selection.

Step 4. Harmony memory update:

In this step, the objective function value is evaluated for the vector x_{new} to determine if the new harmony should be included in the harmony memory. If the new harmony vector is better than the worst harmony in the HM, then the worst harmony is replaced with the new harmony.

Step 5. Repetition of Steps 3 and 4 until the termination criterion is satisfied:

Steps 3 and 4 will be repeated until the stop criterion is satisfied (i.e. maximum number of improvisations).

V. The Use Of Hs For Detecting Distributed Predicates

In this section, we will present an approach based on HS that can be used to detect Possibly: P for any predicate P As we have shown earlier, there are some approaches that can efficiently detect Possibly: P for certain classes of predicates P , like conjunctive predicates. However, there is no efficient approach that can be used to detect Possibly: P for any predicate P. This is due to the fact that the number of global states to be considered in detection is exponential (m^n global states in the worst case, where n is the number of processes and m is the number of local states in each process). Harmony search can be used in this case to provide a powerful general solution in such a case where the search space size to be considered is exponential.

The algorithms developed to detect distributed predicates can be online or offline. Online detection works during the execution of the application, and hence it may change the behavior of the application in unexpected manner. However, it has the advantage of avoiding the need to keep very large trace files as it is the case in offline detection. Offline detection collects the necessary information at runtime and later analyzes it to decide whether a given predicate has been satisfied during the execution or not. Offline detection does not have a strong impact on the behavior of the application under consideration. However, it requires the collection of very large trace files.

In this paper we are adopting the offline approach. In fact the use of harmony search fits more appropriately with the offline approach due to the fact that harmony search operators investigate global states in the whole search space randomly and does not investigate global states in the order in which they may appear during the execution of a given distributed application.

Now we will describe the details of the harmony search algorithm used in our solution. We will start with the representation of a run of a distributed program that the harmony search algorithm can manipulate. We assume that the algorithm to be developed wants to detect the predicate Possibly: P where P is the predicate $x_0 + x_1 + \cdots + x_{n-1} = c$ where x_i is a variable of process P_i and c is a constant.

There is no algorithm presented in the literature to efficiently detect this predicate [1].

We will assume that each process is instrumented to collect at runtime the local states that may affect the desired predicate along with their vector clock timestamp (The vector clock is a very well known technique to assign timestamps to the events of a run of a given distributed program [10], [33]). Consequently, each process P_i will have a trace file of the form

Val_1 , $timestamp_1$

Val_2 , $timestamp_2$

.

.

.

Val_m, $Timestamp_m$

Where Val_j is the value of variable x_i of process P_i in the local state number j. $Timestamp_j$ is the vector clock time stamp of local state j. For example, if we have the run shown in Figure 4 (a) and we want to detect the predicate $x + y + z = 2$, then the trace files of these processes will be as shown in Figure 4 (b). Each trace file contains a list of all local states that may be part of a global state that satisfies the desired predicate in the given run. The local state of a process P_i involves the value of the variables involved in the predicate of interest and the vector clock time stamp of the local state. For example, the first local state of process P_0 is 0, (1, 0, 0). This means that the value of variable x (which is one of the variables involved in the predicate of interest) at this local state is zero and the vector clock time stamp of this local state is (1, 0, 0).

Assuming that we have n processes we will have n trace files. The size of each of them is linear with respect to the number of events executed by each corresponding process. These files will represent the input to the HS algorithm that will be used to detect the above mentioned predicate. The contents of these files can be changed to include additional information if we want to detect other types of predicates. For example, if the predicate of interest involves two or more variables of process P_i , then the values of these variables has to be added to each local state in the trace file of P_i.

Now we will move to the details of the harmony search algorithm itself starting with the representation of the harmony memory (HM) to be used. The HM is a two dimensional array where each row represents a solution of the problem under consideration along with its fitness (quality). In our problem (Distributed predicates detection), each row in the HM represents a global state of the distributed application under consideration. Consequently, each row will have n local states (one from each process) such that all of the local states form a global state of the application. Each global state in the HM is a candidate solution for our problem which is mainly finding a global state that satisfies the predicate to be detected. Moreover, the last element of each row contains the fitness of the solution encoded in that row.

In our example, where we have the run shown in Figure4 (a) and we want to detect the predicate $x + y + z = 2$, the HM can have the form shown in Figure 4 (c) where the size of it is HMS and each row represents a global state along with its fitness. Each element in column i is a local state of process P_i taken from its trace file. For example, the element HM[0][0] shown in Figure 4 (c) is one of the local states of process P_0 taken from its trace file shown in Figure 4 (b).

HS operators may result in solutions that do not represent global states due to the fact that the cuts represented by the resulting chromosomes are not consistent. This problem can be solved by increasing the fitness of such solutions (solutions with small fitness value are better than others with larger fitness value). In our algorithm we will assign the value of 9999999 as the fitness value for any chromosome that does not represent a global state.

Now we will move to the most important step in our algorithm, namely, the design of the fitness function. In fact all of the above steps in the algorithm will be identical for any predicate under consideration. The only difference between the algorithms to detect two different predicates is in the fitness function. This is considered as a strong side in HS. More precisely, HS algorithms are more general than other algorithms in the sense that they can be used to detect any predicate with small modifications on the fitness function. There is no need to develop a tailored algorithm to efficiently detect each type of predicates.

The fitness function has to evaluate each possible solution in the HM and assign to it a fitness value indicating whether the solution is close to the optimal solution or not. In our example, the value assigned by the fitness function to each solution will indicate whether the global state represented by the solution satisfies the predicate of interest or not, and if it does not satisfy the predicate, how close it is to the values that can satisfy the predicate.

The fitness function will first check whether the solution represents a global state or not (consistent cut), if the chromosome does not represent a global state then the fitness value assigned to it will be 9999999 indicating that this solution cannot represent a global state that satisfies the predicate. Otherwise the fitness function will assign a fitness value to the solution according to the following formula:

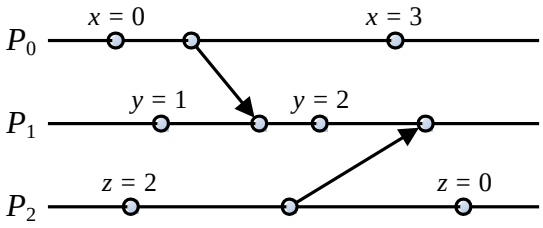
$$Fitness = x_0 + \cdots + x_{n-1} - c$$

For example the fitness of the solution presented in HM[0]

in Figure 4 (c) is $0 + 2 + 2 - 2 = 2$

Consequently, when the solution satisfies the predicate its fitness value will be 0. When the algorithm finds a global state that satisfies the predicate, it will terminate directly since we are looking for the predicates under the possibly modality.

To avoid keep running the algorithm forever in cases there is no global state satisfying the predicate in the run under consideration, we can also fix the maximum number of iterations the HS algorithm can go through. However, it is better to choose a large enough bound on the number of iterations depending on the complexity of the predicate, the number of processes, and the length of the trace files. The larger the number of processes and trace files the larger the number of iterations that need to be considered.



(a)

Trace file of P_0	*Trace file of P_1*	*Trace file of P_2*
0, (1, 0, 0) 3, (3, 0, 0)	1, (0, 1, 0) 2, (2, 3, 0)	2, (0, 0, 1) 0, (0, 0, 3)

(b)

	Harmony Memeory (HM)			*Fitness*
HM[0]	**0, (1, 0, 0)**	**2, (2, 3, 0)**	**2, (0, 0, 1)**	**2**
	∘ ∘ ∘	∘ ∘ ∘	∘ ∘	∘ ∘ ∘
HM[HMS-1]	**3, (3, 0, 0)**	**2, (2, 3, 0)**	**0, (0, 0, 3)**	**3**

(c)

Figure 4. An example to demonstrate the use of HS in detecting distributed predicates.

VI. Implementation And Experimental Results

In this section we will give more details about the implementation of the proposed HS-based distributed predicates detection algorithm. Moreover, we will present some experimental results. We have implemented our algorithm using Java pro- gramming language. We assume that the algorithm to be developed wants to detect the predicate Possibly : P where P is the predicate ($x_0 + \cdots + x_{n-1} = c$) where x_i is a variable of process P_i and c is a constant. We assume that we have n processes. In the previous section we have described the fitness function used in our example. Other parts of the HS algorithm can be implemented in a general manner and can be used in detecting any other predicate. The only thing that has to be changed if we want to detect other predicates is the fitness function.

We have executed our algorithm on a computer with Intel Core 2 Duo CPU, 2.4GHz with 2GB of RAM. We have executed the algorithm on a trace of a distributed application that involves 25, 50, and 75, processes where each process has executed 1000, 2000, 3000 events in each run. We have set the parameters of the HS as follows (HMS = 10, HMCR = 0.9 and PAR = 0.4). The results are summarized in Table 1. For example, given a run that involves 75 processes where each process has executed 3000 events, the algorithm was able to detect the predicate $x_0 + x_1 + \cdots + x_{74} = 67$ after 88457221 iterations and the total time required to detect the predicate was 1687.125 seconds.

The other general approach that can be used to detect any predicate under the possibly modality is to construct the entire state lattice and to test the global states in it one by one until we reach a global state where the predicate of interest is satisfied. This approach requires exploring m^n global states in the worst case where n is the number of processes and m in the number of local states in each process. Obviously, exploring such a large number of states will require much more time than the time required by the HS-based algorithm. For example, a run that involves 20 processes where each process have executed 5 events will have (5^{20}) global states. If we want to examine all the global states in this small run, and assuming that we can examine 10^9 global states per second, then we need 26.49 hours to finish. Consequently, it is clear (See Table 1) that the detection algorithm based on HS is much more powerful.

VII. Conclusions And Future Work

In this paper, we have developed an efficient distributed predicates detection algorithm based on harmony search. The proposed algorithm can be used to detect distributed predicates under the possibly modality. Distributed predicates detection (also referred to as runtime verification) is an effective technique to reason about a particular implementation of a given distributed application. Consequently, the results presented in this paper can be exploited in developing more dependable distributed applications.

Genetic algorithms (GAs) can be used effectively to solve problems with exponential size search space as it is the case in distributed predicates' detection. In one of our current research efforts we are trying to investigate the effects of using the mutation operator of GAs in HS to solve certain optimization problems.

Table I
THE RESULTS OF SEVERAL EXPERIMENTS.

Number of Processes	Number of events executed by each process	Time spent to detect the predicate/ seconds	Number of Iterations
25	1000	0.031	72
	2000	0.047	282
	3000	0.062	587
50	1000	1.234	92580
	2000	1.5	117323
	3000	2.032	159467
75	1000	830.454	44732576
	2000	1504.641	80262355
	3000	1687.125	88457221

One possible avenue for the continuation of the work presented in this paper is to consider the effects of using the mutation operators of GAs on the performance of the HS algorithm developed to detect distributed predicates. Moreover, we can develop a detecting approached based completely on GAs and compare it with the approach developed based on HS. Another important avenue for future work is improving the fitness function. In fact, if we can find a more powerful fitness function, then the performance of the proposed algorithm will certainly be improved.

REFERENCES

[1] Vijay K. Garg, Elements of distributed computing, John Wiley & Sons, Inc., New York, NY, USA, 2002.

[2] G. Dumais and H.F. Li, "Distributed predicate detection in series-parallel systems," IEEE Transactions on Parallel and Distributed Systems, vol. 13, no. 4, pp. 373 –387, apr 2002.

[3] Felix C. Freiling and Arshad Jhumka, "Global predicate detection in distributed systems with small faults," in Pro- ceedings of the 9th international conference on Stabilization, safety, and security of distributed systems, Berlin, Heidelberg,2007, SSS'07, pp. 296–310, Springer-Verlag.

[4] Chunbo Chu and M. Brockmeyer, "Predicate detection modality and semantics in three partially synchronous mod- els," in the Seventh IEEE/ACIS International Conference on Computer and Information Science, may 2008, pp.444–450.

[5] Vijay K. Garg and Brian Waldecker, "Detection of weak unstable predicates in distributed programs," IEEE Trans. Parallel Distrib. Syst., vol. 5, no. 3, pp. 299–307, 1994.

[6] Vijay Garg and Brian Waldecker, "Detection of strong unstable predicates in distributed programs," IEEE Trans. Parallel Distrib. Syst., vol. 7, no. 12, pp. 1323–1333, 1996.

[7] Eslam Al Maghayreh, Simplifying Runtime Verification of Distributed Programs: Ameliorating the State Space Explosion Problem, VDM Verlag, 2010.

[8] Craig M. Chase and Vijay K. Garg, "Detection of global predicates: Techniques and their limitations," Distributed Computing, vol. 11, no. 4, pp. 191–201, 1998.

[9] Leslie Lamport, "Time, clocks, and the ordering of events in a distributed system," Commun. ACM, vol. 21, no. 7, pp. 558–565, 1978.

[10] Friedemann Mattern, "Virtual Time and Global States of Distributed Systems," in Proceedings of the International Workshop on Parallel and Distributed Algorithms, Château de Bonas, France, October 1989, pp. 215–226.

[11] Robert Cooper and Keith Marzullo, "Consistent detection of global predicates," SIGPLAN Not., vol. 26, no. 12, pp. 167– 174, 1991.

[12] Roland Jegou, Raoul Medina, and Lhouari Nourine, "Linear space algorithm for on-line detection of global predicates.," in Proceedings of the International Workshop on Structures in Concurrency Theory (STRICT), Berlin, 1995, pp. 175–189.

[13] K. Mani Chandy and Leslie Lamport, "Distributed snapshots: Determining global states of distributed systems,"ACM Trans. Comput. Syst., vol. 3, no. 1, pp. 63–75, 1985.

[14] Luc Bougé, "Repeated snapshots in distributed systems with synchronous communications and their implementation in CSP," Theor. Comput. Sci., vol. 49, pp. 145–169, 1987.

[15] Madalene Spezialetti and Phil Kearns, "Efficient distributed snapshots," in ICDCS, 1986, pp. 382–388.

[16] H. F. Li, Juergen Rilling, and Dhrubajyoti Goswami, "Granularity-driven dynamic predicate slicing algorithms for message passing systems," Automated Software Engg., vol. 11, no. 1, pp. 63–89, 2004.

[17] Alper Sen and Vijay K. Garg, "Detecting temporal logic pred- icates in distributed programs using computation slicing." in OPODIS, 2003, pp. 171–183.

[18] Vijay K. Garg and Neeraj Mittal, "On slicing a distributed computation," in ICDCS '01: Proceedings of the The 21st nternational Conference on Distributed Computing Systems, 2001, p. 322.

[19] Neeraj Mittal and Vijay K. Garg, "Computation slicing: Techniques and theory," in DISC '01: Proceedings of the 15th International Conference on Distributed Computing, London, UK, 2001, pp. 78–92, Springer-Verlag.

[20] L. Lamport, "A theorem on atomicity in distributed algo- rithms," Distributed Computing, vol. 4, no. 2, pp. 59–68, 1990.

[21] Eslam Al Maghayreh, "Block-based atomicity to simplify the verification of distributed applications," in 24th Cana- dian Conference on Electrical and Computer Engineering (CCECE), may 2011, pp. 887–891.

[22] H. F. Li, Eslam Al Maghayreh, and D. Goswami, "Detecting atomicity errors in message passing programs," in PDCAT'07: Proceedings of the Eighth International Conference on Parallel and Distributed Computing, Applications and Technologies. 2007, pp. 193–200, IEEE Computer Society.

[23] H. F. Li, Eslam Al Maghayreh, and D. Goswami, "Using atoms to simplify distributed programs checking," in DASC'07: Proceedings of the Third IEEE International Symposium on Dependable, Autonomic and Secure Computing, 2007, pp.75–83.

[24] Zong Woo Geem, Joong-Hoon Kim, and G. V. Loganathan, "A new heuristic optimization algorithm: Harmony search," Simulation, vol. 76, no. 2, pp. 60–68, 2001.

[25] Zong Geem, "Novel derivative of harmony search algorithm for discrete design variables," Applied Mathematics and Computation, vol. 199, no. 1, pp. 223–230, 2008.

[26] Kang S. Lee and Zong W. Geem, "A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory and practice," Computer Methods in Applied Mechanics and Engineering, vol. 194, no. 36-38, pp. 3902–3933, Sept. 2005.

[27] Zong Woo Geem, "Global optimization using harmony search: Theoretical foundations and applications," in Foun- dations of Computational Intelligence (3), pp. 57–73. 2009.

[28] Zong Geem, "Harmony search algorithm for solving sudoku," in Knowledge-Based Intelligent Information and Engineering Systems (KES), 2007, pp. 371–378.

[29] Zong Geem, "Harmony search for multiple dam scheduling," in Encyclopedia of Artificial Intelligence, pp. 803–807. 2009.

[30] Osama Moh'd Alia, Mohammed Azmi Al-Betar, Mandava Rajeswari, and Ahamad Tajudin Khader, "Data clustering using harmony search

algorithm," in Proceedings of Second International Conference Swarm, Evolutionary, and Memetic Computing (SEMCCO 2), 2011, pp. 79–88.

[31] Mohammed Azmi Al-Betar and Ahamad Tajudin Khader, "A harmony search algorithm for university course timetabling," Annals OR, vol. 194, no. 1, pp. 3–31, 2012.

[32] Zong Woo Geem, Kang Seok Lee, and Chung-Li Tseng, "Harmony search for structural design," in Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2005), 2005, pp. 651–652.

[33] Colin Fidge, "Timestamps in Message-Passing Systems that Preserve the Partial Ordering," in Proceedings of the 11th Australian Computer Science Conference, 1988, pp. 56–66.

A Novel Technique for Glitch and Leakage Power Reduction in CMOS VLSI Circuits

Pushpa Saini
M.E. Student, Department of Electronics and Communication Engineering
NITTTR, Chandigarh, India

Rajesh Mehra
Associate Professor, Department of Electronics and Communication Engineering
NITTTR, Chandigarh, India

***Abstract*—Leakage power has become a serious concern in nanometer CMOS technologies. Dynamic and leakage power both are the main contributors to the total power consumption. In the past, the dynamic power has dominated the total power dissipation of CMOS devices. However, with the continuous trend of technology scaling, leakage power is becoming a main contributor to power consumption. In this paper, a technique has been proposed which will reduce simultaneously both glitch and leakage power. The results are simulated in Microwind3.1 in 90nm and 250 nm technology at room temperature.**

Keywords-*Dynamic power; Leakage power; Multi-threshold; Variable body biasing; Glitch.*

I. INTRODUCTION

The development of digital integrated circuits is challenged by higher power consumption. The combination of higher clock speeds, greater functional integration, and smaller process geometries has contributed to significant growth in power density. Scaling improves transistor density and functionality on a chip. Scaling helps to increase speed and frequency of operation and hence higher performance. As voltages scale downward with the geometries threshold voltages must also decrease to gain the performance advantages of the new technology, but leakage current increases exponentially. Thinner gate oxides have led to an increase in gate leakage current.

Today leakage power has become an increasingly important issue in processor hardware and software design. With the main component of leakage, the sub-threshold current, exponentially increasing with decreasing device dimensions, leakage commands an ever increasing share in the processor power consumption. In 65 nm and below technologies, leakage accounts for 30-40% of processor power.

According to the International Technology Roadmap for Semiconductors (ITRS), leakage power dissipation may eventually dominate total power consumption as technology feature sizes shrink. While there are several process technology and circuit-level solutions to reduce leakage in processors, in this paper a novel approaches for reducing both leakage and dynamic power with minimum possible area and delay tradeoff are proposed.

For the most recent CMOS feature sizes (e.g., 90nm and 65nm), leakage power dissipation has become an overriding concern for VLSI circuit designers. For deep-submicron processes, supply voltages and threshold voltages for MOS transistors are greatly reduced. This to an extent reduces the dynamic (switching) power dissipation. However, the subthreshold leakage current increases exponentially thereby increasing static power dissipation [1].

Power consumption of CMOS consists of dynamic and static components. Dynamic power is consumed when transistors are switching, and static power is consumed regardless of transistor switching. Dynamic power consumption was previously (at 0.18µ technology and above) the single largest concern for low-power chip designers since dynamic power accounted for 90% or more of the total chip power. Therefore, many previously proposed techniques, such as voltage and frequency scaling, focused on dynamic power reduction. However, as the feature size shrinks, e.g., to 0.09µ and 0.065µ, static power has become a great challenge for current and future technologies.

Modern digital circuits consist of logic gates implemented in the complementary metal oxide semiconductor (CMOS) technology. Power consumption has two components: Dynamic Power and Leakage power [2]. Dynamic and leakage power both are the main contributors to the total power consumption. Dynamic power includes both switching power and short circuit power. Spurious transitions (also called glitches) in combinational CMOS logic are a well-known source of unnecessary power dissipation. Reducing glitch power is a highly desirable target [3]. The dynamic power cannot be eliminated completely, because it is caused by the computing activity. It can, however, be reduced by circuit design techniques.

Static power refers to the power dissipation which results from the current leakage produced by CMOS transistor parasitic. Traditionally static power has been overshadowed by dynamic power consumption, but as transistor sizes continue to shrink, static power may overtake dynamic power consumption To alleviate the rising significance of static power in digital systems, static power reduction technique shave been developed like transistor stacking, dual threshold voltage , MTCMOS etc. Some of these techniques are state saving and some are state destructive techniques. For example: Sleep transistor is a state destructive technique. Despite the rising significance of static power in CMOS circuits, the dynamic power is still the major contributor to power consumption. Dynamic power is mostly consumed by glitches which are the unwanted transitions and need to be eliminated.

Glitch and leakage power both are the main contributors to the power consumption and needs to be reduced.

II. Power Dissipation Factors

In CMOS, power consumption consists of leakage power and dynamic power Dynamic power includes both switching power and short circuit power. Switching power is consumed when the transistors are in active mode and short circuit power is consume when a pull-up and pull-down network are on turning on and off. For 0.18u and above leakage power is small compared to dynamic power but 0.13u and below leakage power is dominant. Dynamic power dissipation is proportional to the square of the supply voltage. In deep sub-micron processes, supply voltages and threshold voltages for MOS transistors are greatly reduced. This, to an extent, reduces the dynamic power dissipation [4].

Static power dissipation is the power dissipation due to leakage currents which flow through a transistor when no transactions occur and the transistor is in a steady state. Leakage power depends on gate length and oxide thickness. It varies exponentially with threshold voltage and other parameters. Reduction of supply voltages and threshold voltages for MOS transistors, which helps to reduce dynamic power dissipation, becomes disadvantageous in this case. The subthreshold leakage current increases exponentially, thereby increasing static power dissipation.

The leakage current of a transistor is mainly the result of reverse biased PN junction leakage and Sub threshold leakage. Compared to the subthreshold leakage, the reverse bias PN junction leakage can be ignored. The Subthreshold conduction or the subthreshold leakage or the subthreshold drain current is the current that flows between the source and drain of a MOSFET when the transistor is in subthreshold region, or weak-inversion region, that is, for gate-to-source voltages below the threshold voltage [5].

It is given by:

$$I_{sub} = I_{so} \exp\left(\frac{V_{gs} - V_{th}}{V_T}\right)\left(1 - \exp\left(\frac{-V_{ds}}{V_T}\right)\right) \quad (1)$$

$$I_{s0} = \mu_0 C_{ox} \frac{W_{eff}}{L_{eff}} \quad (2)$$

where μ_0 is the zero bias electron mobility, n is the subthreshold slope coefficient, V_{gs} and V_{ds} are the gate to source voltage and drain-to-source voltage, respectively, V_T is the thermal voltage, V_{th} is the threshold voltage, C_{ox} is the oxide capacitance per unit area, and W_{eff} and L_{eff} are the effective channel width and length, respectively. Due to the exponential relation between V_{th} and I_{sub}, an increase in V_{th} sharply reduces the subthreshold current.

A. Leakage Current Reduction

Reduction in threshold voltage results in the increase in sub-threshold leakage current. One of challenge with technology scaling is the rapid increase in subthreshold leakage power due to V_t reduction. In such a system it becomes crucial to identify techniques to reduce this leakage power component. The development of digital integrated circuits is challenged by higher power consumption [6].

Leakage current is a primary concern for low-power, high-performance digital CMOS circuits. The exponential increase in the leakage component of the total chip power can be attributed to threshold voltage scaling, which is essential to maintain high performance in active mode, since supply voltages are scaled. Numerous design techniques have been proposed to reduce standby leakage in digital circuits. Leakage power has become a serious concern in nanometer CMOS technologies, and power-gating has shown to offer a viable solution to the problem with a small penalty in performance [7].

Devices which are operated on battery are either idle (Standby) or Active mode. Leakage power can be divided in to two categories based on these two modes [8]:

1) Leakage Control in Standby Mode: Techniques like Power gating and super cutoff CMOS are used for leakage reduction in standby mode. In these techniques, circuit is cutoff from the supply rails, when it is in idle state.

2) Leakage Control in Active Mode: Techniques like forced stacking and sleepy stack can be used during the run time or active mode for leakage current reduction.

Leakage is becoming comparable to dynamic switching power with the continuous scaling down of CMOS technology. To reduce leakage power, many techniques have been proposed, including dual-Vth, multi-Vth, optimal standby input vector selection, transistor stacking, and body bias.

Multiple thresholds can be used to deal with the leakage problem in low-voltage high-performance CMOS circuits. The dual-V_{th} assignment is an efficient technique for decreasing leakage power. In this method, each cell in the standard cell library has two versions, low Vthand high Vth. Gates with low Vthare fast, but have high subthreshold leakage, whereas gates with high Vthare slower but have much reduced subthreshold leakage. The generation, distribution, and dissipation of power are at the forefront of current problems faced by the integrated circuit industry. The application of aggressive circuit design techniques which only focus on enhancing circuit speed without considering power is no longer an acceptable approach in most high complexity [9]. Already existing methods like stack, sleepy stack, and sleep transistor are shown in Fig. 1-3.

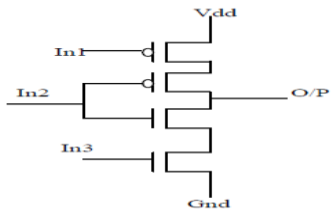


Figure 1. Sleep Transistor.

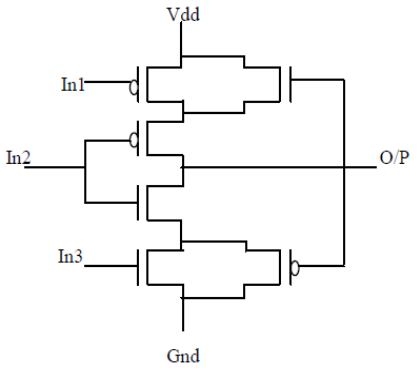


Figure 2.Sleepy Keeper.

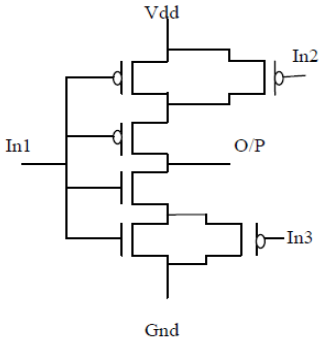


Figure 3.Sleepy Stack.

Circuit optimization provides low power and high performance. Circuit optimization can be obtained through simultaneous gate sizing and threshold voltage (Vt) assignment [10,11]. The sleep transistors areturned off when the logic circuits are not in use. By isolating the logic networks using sleeptransistors, the sleep transistor technique dramatically reduces leakage power during sleepmode.Sleep transistor method provides good reduction in leakage power, but it is a state destructive technique. It is shown in Fig. 1.

State -destructive techniques cut off transistor (pull-up or pull-down or both) networks from supply voltage or ground using sleep transistors. Both dynamic and leakage power reductions can be achieved through threshold voltage adjustment [12]. Sleepy keeper technique shown in Fig. 2 uses the traditional sleep transistors with two additional transistors to save state during sleep mode. Dual threshold voltages can also be applied in the sleepy keeper approach to reduce subthreshold leakage current [13].The sleepy stack approach combines the sleep and stack approaches. The stack approach uses a stack effect by breaking down an existing transistor into two half size transistors [14, 15, 16]. The sleepy stack technique divides existing transistors into two half size transistors like the stack approach. Then sleep transistors are added in parallel to one of the divided transistors. Fig. 3 shows its structure.

B. CMOS Glitch Elimination

One of the major factors contributing to the power dissipation in CMOS digital circuits is the switching activity. Dynamic power comprises of two parts: Logic switching power and glitch power. Whenever a logic gate changes state, power is consumed. The state change can be due to the essential logic value changes as well as due to glitches. Every signal transition consumes a finite amount of energy. For the correct functioning of a logic circuit, every signal needs to transition at most one time in one clock cycle. But in reality, the gate outputs transition more than once and these unnecessary transitions are called glitches. These transitions consume energy and are quite unnecessary for the correct functioning of the circuit.

Because switching power consumed by the gate is directly proportional to the number of output transitions, glitches reportedly account for 20%– 70% dynamic power. Delay elements are components inserted into a digital circuit that do not alter the signal value, but deliver the same waveform at the output with some extra delay. Different delay elements can be used to insert delay at the inputs of gate. By inserting these delay elements glitches can be eliminated. A buffer is the simplest of the delay elements. Insertion of the buffer as the delay element is one of the way to remove glitches or unwanted transitions. Buffer as delay elements are simple and reliable, but their problem is increased dynamic power. NMOS, Transmission Gate, Cascaded Inverters are some of the other delay elements.

A combinational circuit is minimum transient energy design, i.e., there is no glitch at the output of any gate, if the difference of the signal arrival times at every gate's inputs remains smaller than the inertial delay of the gate. Hazard filtering, when used alone for glitch elimination, can increase the overall input to output delay. Path balancing does not increase the delay but requires insertion of delay elements. A combination of the two procedures can give an optimum design.

III. PROPOSED MODEL

Low power has emerged as a principal theme in today's electronics industry. The need for low power has caused a major paradigm shift, where power dissipation has become as important a consideration as performance and area. Two components determine the power consumption in a CMOS circuit: Static and Dynamic Power. Static (Leakage) power: includes sub-threshold leakage, drain junction leakage and gate leakage due to tunneling.

Among these, subthreshold leakage is the most prominent one. Dynamic power: Includes charging and discharging (switching) power and short circuit power. In Dynamic power, power consumption due to switching activity is more prominent. It can be concluded from the above discussion so far that glitch and leakage power both are the main contributors to the power consumption.

The existing leakage reduction techniques like sleep transistor, sleepy keeper, stack etc. are having the drawbacks like: increased delay, area etc. and the buffer used as delay elements for elimination of glitches also has the drawbacks of large area overhead and increases the number of transitions in the output. Therefore, in this section new approach has been proposed keeping in mind all the drawbacks mentioned above, which will simultaneously reduce both glitch and leakage power.

A novel technique has been proposed in this section, which will reduce both glitch and leakage power in CMOS VLSI circuits. The new technique is Sleep Variable body biasing with transmission gate. The circuit diagram of unoptimized circuit 1 and optimized circuit 1 is shown in the Fig.4-5.

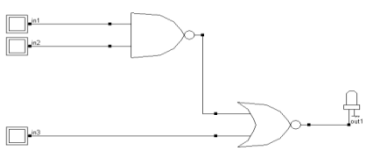


Figure 4.Unoptimized Circuit 1.

This proposed design includes variable body biasing technique along with sleep insertion technique. Sleep transistors are crucial part in any low leakage power design. The source of one of the sleep transistor is connected to the body of other PMOS sleep transistor for having body biasing effect. So, leakage reduction in this technique occurs in two ways. Firstly, the sleep transistor effect and secondly, the variable body biasing effect. This technique uses aspect ratio W/L=3 for NMOS transistor and W/L=6 for PMOS transistor. Due to the minimum aspect ratio the sub-threshold current reduces.

Since the sources of the NMOS sleep transistor is connected to the body of PMOS transistor as shown in Fig. 5, the threshold voltage of the sleep transistors increases due to the body biasing effect during sleep mode. This increase of threshold voltage of the transistors reduces the leakage current. That's why the static power consumption also lowers.

The variable biasing will be useful in reducing leakage power. Sleep transistor method provides good reduction in leakage power in idle mode, but it is a state destructive technique.

Stacking approach is also utilized here to some extent to retain the state in active mode. Variable body biasing will be useful in increasing threshold voltage to reduce leakage current.

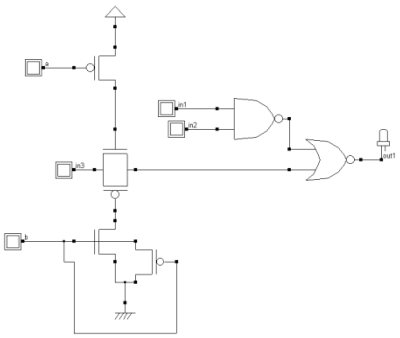


Figure 5.Optimized Circuit 1.

For the reduction of glitch power a transmission gate is also included. The transmission gate is used as a delay element for the elimination of the glitches. The transmission gate has a less area overhead as compared to other delay elements.

The technique has been used on non-critical paths to reduce glitches.

Consider another circuit diagramunoptimized circuit 2 shown in Fig. 6 [2]. The output of this circuit also has glitches, which is a waste of energy. Glitches are occurring because there is difference in the signal arrival times at the inputs of gate. The proposed technique is also applied in this circuit. This circuit is simulated in both 250nm and 90nm technology.

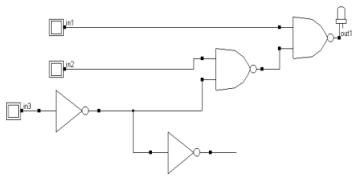


Figure 6.Unoptimized Circuit 2 [2].

The proposed technique is applied in the unoptimized circuit shown in Fig. 6. The optimized circuit is shown in Fig. 7.Transmission gate used here in the proposed technique is useful for eliminating glitches present in the output of unoptimized circuit.

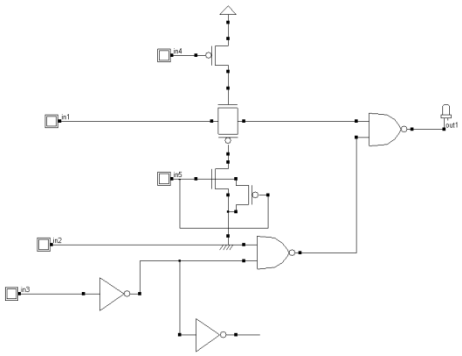


Figure 7.Optimized Circuit 2.

IV. Results And Conclusions

In this section, simulations of the proposed methods Comparisons between optimized and unoptimized circuit are shown in tabular form. Simulations are obtained in Microwind Tool. First step in obtaining the simulations is to compile the Verilog file in Microwind 3.1.

Verilog file is created from the circuit diagram, which is designed in the schematic. The Verilog file is now compiled in Microwind 3.1. After the compilation of Verilog file, the layout for the circuit diagram drawn in schematic will be generated in Microwind. After that simulations are performed on the layout generated using Verilog files. The results are simulated at room temperature.

Simulations of circuits given in Fig. 4-7 are shown below in Fig.8-13.Simulations shown in these figures include the waveform of Voltage vs. Time and Voltage vs. Current. Simulations for unoptimized and optimized circuit 2 shown in Fig. 6-7 are given for both 250 and 90 nm technology.

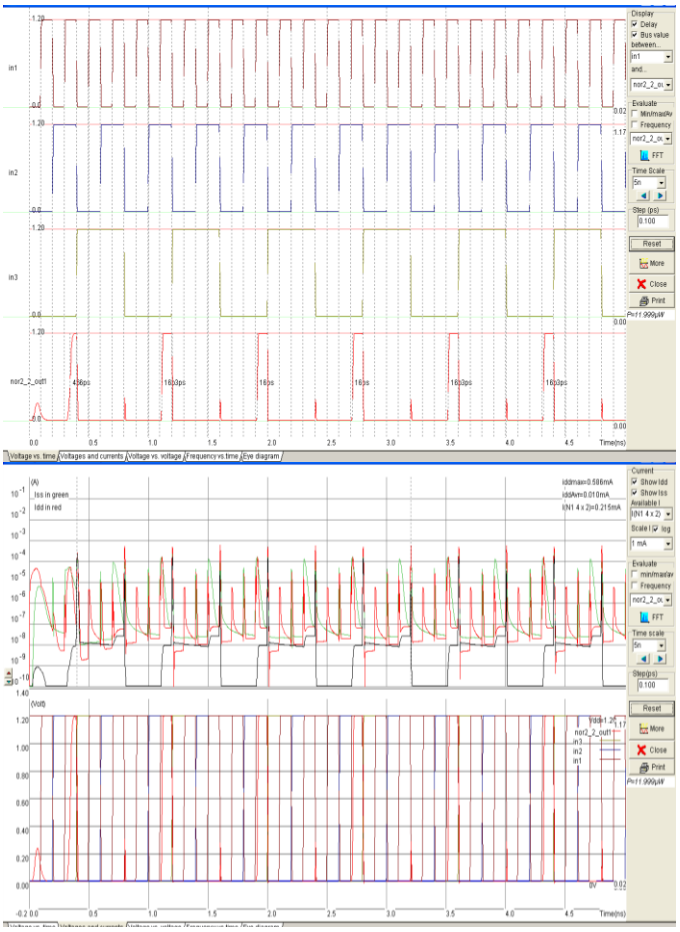


Figure 8.Simulations of Unoptimized Circuit 1(90 nm).

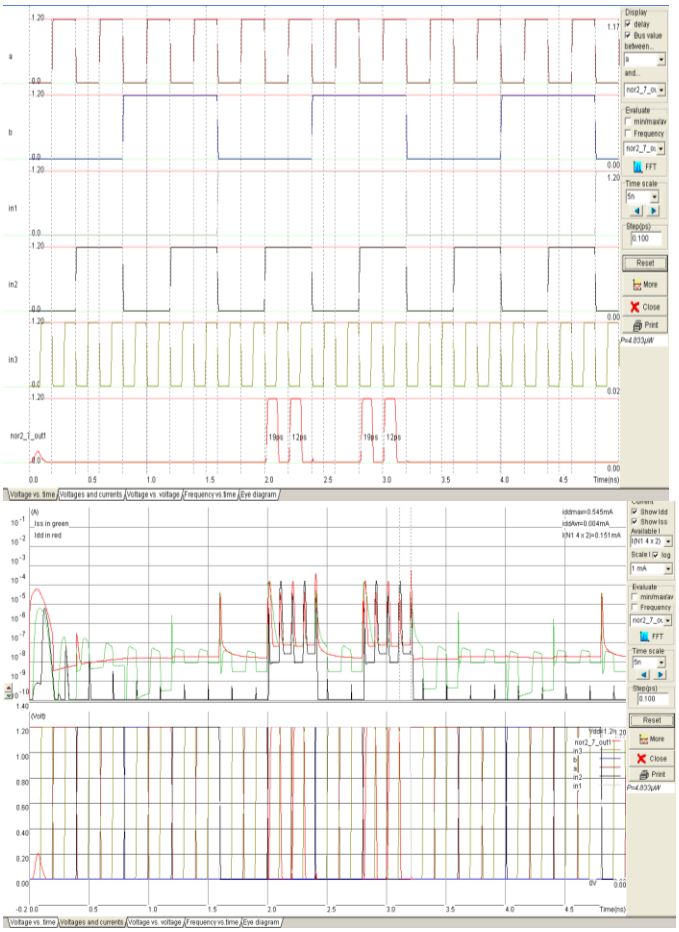


Figure 9.Simulations of Optimized Circuit 1(90 nm).

A. *Results*

Simulations of an unoptimized circuit are shown in Fig. 8. It can be observed from the simulations that glitches are present in the output, which are unwanted transitions and need to be eliminated or reduced. No method to reduce leakage is present in this circuit; due to this leakage current of about 0.215 mA is present as can be observed from waveforms. Glitches present in the O/P and leakage current are major reason here for power consumption.

Simulations of an optimized circuit are shown in Fig. 9. It can be observed from the simulations that glitches are completely eliminated as well as there is a reduction of about 29% in leakage current as observed from the simulations. Because of reduction in leakage current and elimination of glitches, there is a considerable reduction in power consumption. Delay in an optimized circuit is also less as compared to unoptimized circuit.

Simulations for Fig. 6 are shown in Fig. 10 for 250 nm technology. As it can be observed from simulations, glitches are present in the output. Simulations of Optimized circuit 2 shown in Fig. 7 are shown in Fig. 11 for 250 nm technology. Glitches are completely eliminated from optimized circuit's output and considerable reduction in power is also obtained here.

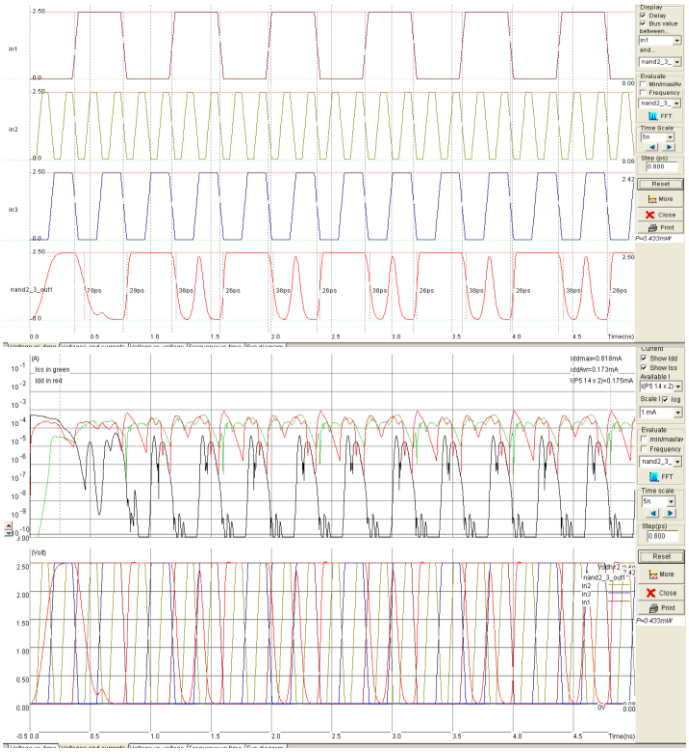


Figure 10.Simulations of Unoptimized Circuit 2 (250nm).

Unoptimized Circuit shown in Fig. 6 is also implemented in 90nm technology and its simulations are shown in Fig. 12.Unwanted transitions are also present here. Due to scaling of technology, the leakage current is also present. Power consumption is due to both the leakage current and these unwanted transitions.

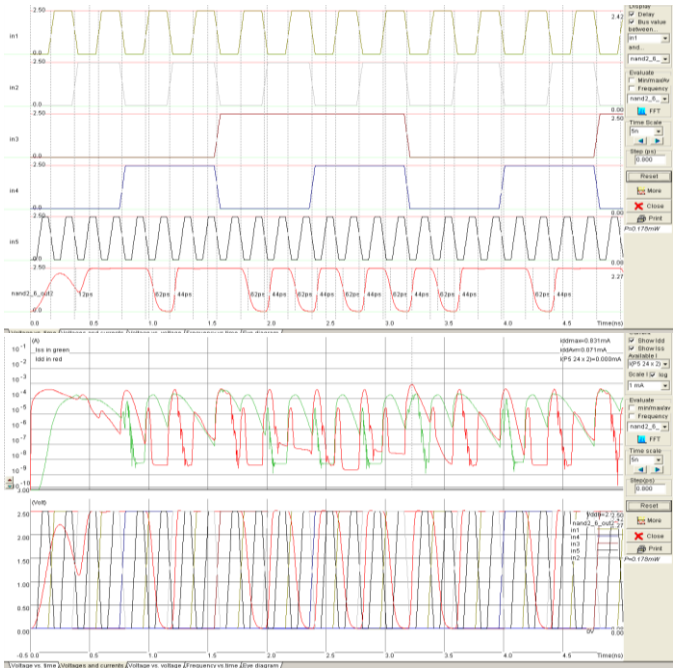


Figure 11.Simulations of Optimized Circuit 2(250 nm).

Simulations of optimized circuit 2 (90 nm) in Fig. 7 are shown in Fig. 13. There is considerable reduction in leakage current as observed from simulations as well as power. But there is little increase in delay of about 1.5ps.

B. Conclusion

Scaling down of the technology has led to increase in leakage current. Nowadays, a leakage power has become more dominant as compared to Dynamic power. But, Dynamic Power consumption due to glitches can't be neglected.

Therefore, in this paper, the efficient technique has been proposed for reducing glitch and leakage power reduction in CMOS VLSI Circuits. The proposed method results in ultra low power consumption.

Two optimized circuits are giving good results in terms of power delay, energy and leakage current as compared to unoptimized circuits. Reduction of about 59.7% is obtained in power and in energy it is 85.28% for optimized circuit shown in Fig. 5 as compared to unoptimized circuit given in Fig. 4.The comparison is shown in Table I given below. The results are simulated using Microwind 3.1 tool in 90nm technology at room temperature for the circuits shown in Fig. 4-5.

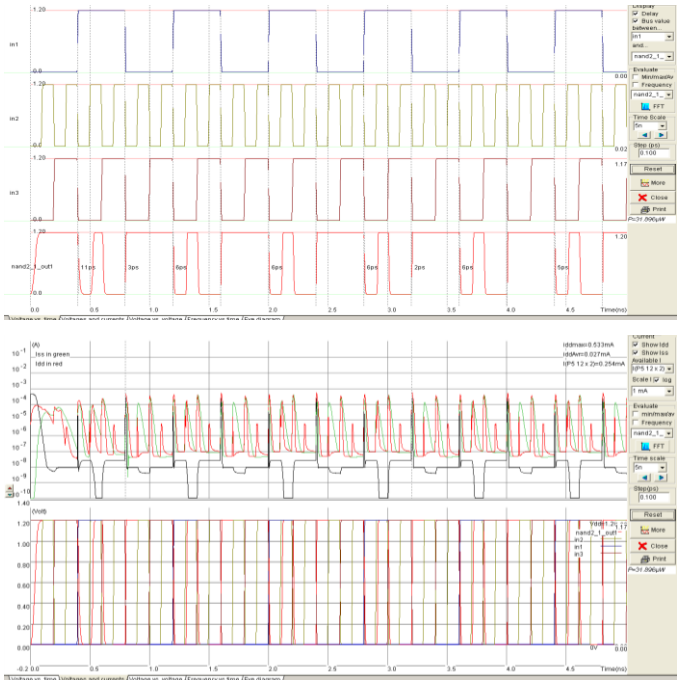
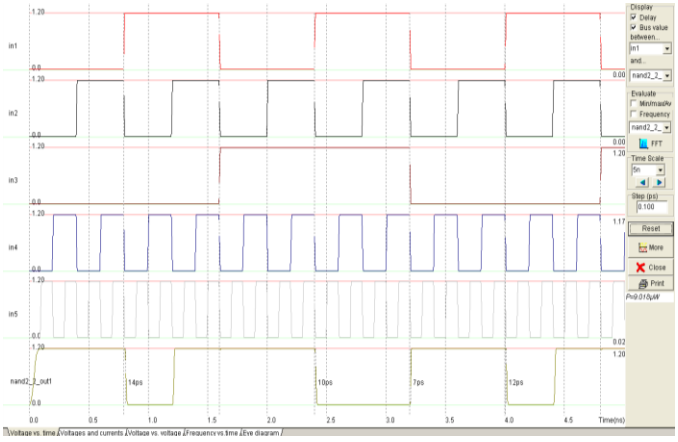


Figure 12.Simulations of Unoptimized Circuit 2 (90 nm).

Circuits shown in Fig. 6-7 are simulated in Microwind 3.1 for both 250 nm and 90 nm technology. Comparison table given in Table II has shown the comparison between unoptimized circuit and optimized circuit 2 for 250 nm technology. An optimized circuit 2 (250 nm) has about 67% energy and 58.8% power reduction as compared to unoptimized circuit 2 (250 nm).

Table III is showing the comparison for 90nm technology.There is about 64.02% energy and 71.72% power reduction in an optimized circuit 2 (90 nm) over unoptimized circuit 2 (90 nm). As it can be observed from results shown in comparison tables, optimized circuits are more energy and power efficient as compared to unoptimized circuits. Glitches are also completely eliminated from outputs of optimized circuits.



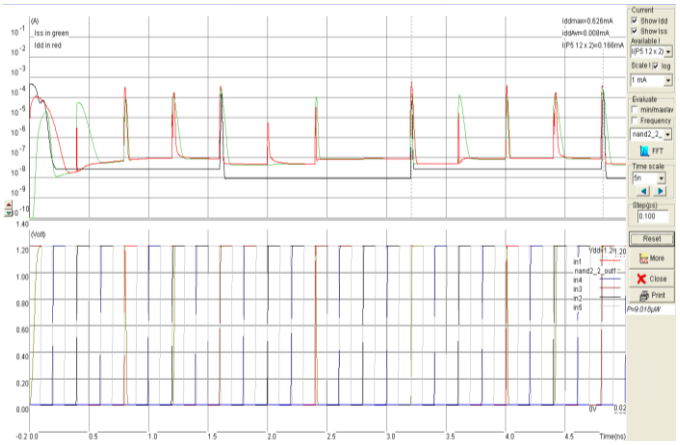


Figure 13.Simulations of Optimized Circuit 2 (90 nm).

TABLE I. COMPARISON BETWEEN UNOPTIMIZED AND OPTIMIZED CIRCUIT 1(90 nm)

Circuit / Parameter	*Unoptimized Circuit 1*	*Optimized Circuit 1*
Power(µW)	11.999	4.833
Delay(ps)	26	9.5
Energy(aJ)	311.974	45.9135
Current(mA)	0.586	0.545
Leakage Current(mA)	0.215	0.151

TABLE II COMPARISON BETWEENNOPTIMIZED AND OPTIMIZED CIRCUIT 2 (250nm)

Circuit / Parameter	*Unoptimized Circuit 2*	*Optimized Circuit 2*
Power(mW)	0.433	0.178
Delay(ps)	39	31
Energy(fJ)	16.887	5.518
Current(mA)	0.818	0.831
Leakage Current(mA)	0.175	0.000

TABLE III COMPARISON BETWEEN UNOPTIMIZED AND OPTIMIZED CIRCUIT 2 (90 nm)

Circuit / Parameter	*Unoptimized Circuit 2*	*Optimized Circuit 2*
Power(µW)	31.896	9.018
Delay(ps)	5.5	7
Energy(aJ)	175.428	63.126
Current(mA)	0.533	0.626
Leakage Current(mA)	0.254	0.166

ACKNOWLEDGMENT

The authors would like to thank Dr. Samir Kumar Das, Director, NITTTR, Chandigarh for constant encouragement and support during this research work. The authors would also like to express their sincere thanks and deep sense of gratitude to Dr. SBL Sachan, Professor and Head, Electronics & Communication Department, NITTTR, Chandigarh for their constant inspirations, guidance and helpful suggestions throughout this research work.

REFERENCES

[1] Jae Woong Chun and C. Y. Roger Chen, "A Novel Leakage Power Reduction Technique for CMOS Circuit Design", International Conference on SoC Design Conference (ISOCC), pp. 119-122 , IEEE 2010

[2] Tezaswi Raja, VishwaniD.Agrawal and Michael L. Bushnell "Variable Input Delay CMOS Logic for Low Power Design", IEEE Transactions on Very Large Scale Integration (VLSI) System, Vol. 17, Issue: 10, pp. 1534-1545, 2009

[3] SarveshBhardwaj and SarmaVrudhula, "Leakage Minimization of Digital Circuits Using Gate Sizing in the Presence of Process Variations", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 27, Issue: 3, pp. 445-455, March 2008.

[4] Anup K. Sultania, Dennis Sylvester, and Sachin S. Sapatnekar, "Gate Oxide Leakage and Delay Tradeoffs for Dual-Tox Circuits", IEEE Transactions on Very Large Scale Integration (VLSI) systems, Vol. 13, Issue: 12, pp. 1362-1375, December 2005.

[5] Yuanlin Lu and Vishwani D. Agrawal "CMOS Leakage and Glitch Minimization for Power Performance Tradeoff" IEEE Journal of Low Power Electronics, Vol. 2, pp. 1-10, 2006.

[6] M. S. Islam, M. Sultana Nasrin, Nuzhat Mansur and NailaTasneem, "Dual Stack Method: A Novel Approach to Low Leakage and Speed Power Product VLSI Design", 6th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, pp. 18-20, IEEE December 2010.

[7] AshokaSanthanur, Luca Benini, "Row–Based Power– Gating: A Novel Sleep Transistor Insertion Methodology for Leakage Power Optimization in Nanometer CMOS Circuits", IEEE Transactions on VLSI Systems, Vol. 19, Issue: 3, pp. 469-482, March 2011.

[8] Fallah, F., and Pedram, M., Standby and Active Leakage Current Control and Minimization in CMOS VLSI Circuits. IEICE Transactions on Electronics, Special Section on Low-Power LSI and Low-Power IP E88-C, 4 (April 2005), 509-519.

[9] Salendra. Govindarajulu, "Low Power, Reduced Dynamic Voltage Swing Domino Logic Circuits" Indian Journal of Computer Science and Engineering, Vol. 1, No. 2, pp. 74-81.

[10] Heung Jun Jeon, Yong-Bin Kim, and Minsu Choi, "Standby Leakage Power Reduction Technique for Nanoscale CMOS VLSI Systems", IEEE Transactions on Instrumentation and Measurement, Vol. 59, No. 5, pp. 1127-1133, May 2010.

[11] Shuzhe Zhou, Hailong Yao, Qiang Zhou, "Minimization of Circuit Delay and Power through Gate Sizing and Threshold Voltage Assignment", IEEE Computer Society Annual Symposium on VLSI, pp. 212-217, 2011.

[12] Philippe Matherat, MariemSlimani "Multiple Threshold Voltage for Glitch Power Reduction", Faible Tension FaibleConsommation (FTFC), pp. 67-70, IEEE 2011.

[13] Se Hun Kim, Vincent J. Mooney III, "Sleepy Keeper: a New Approach to Low-leakage Power VLSI Design", 2006 IFIP International Conference on Very Large Scale Integration, Oct. 2006, pp.367 – 372.

[14] Jun Cheol Park, Vincent J. Mooney III, and Philipp Pfeiffenberger , "Sleepy Stack Reduction of Leakage Power", IEEE Transactions on Very Large Scale Integration (VLSI) Systems,Vol.14, No.11, pp. 1250–1263, 2006.

[15] S. G. Narendra and A. P. Chandrakasan, Leakage in Nanometer CMOS Technologies. Berlin, Germany: Springer-Verlag, 2006, pp. 21–40.

[16] K. K. Kim and Y.-B. Kim, "A novel adaptive design methodology for minimum leakage power considering PVT variations on nanoscale VLSI systems," IEEE Trans. Very Large Scale Integration (VLSI), Vol. 17, No. 4, pp. 517–528, Apr. 2009.

AUTHOR'S PROFILE

PushpaSaini: Mrs. PushpaSaini is currently pursuing M.E. (E&C) from National Institute of Technical Teachers' Training & Research, Chandigarh, Panjab University. She has completed her B.Tech. from MMEC, Kurukshetra University, India.

Rajesh Mehra: Mr. Rajesh Mehra is currently Associate Professor at National Institute ofTechnical Teachers' Training & Research, Chandigarh, India. He is pursuing his PhD from Panjab University,Chandigarh, India. He has completed his M.E. from NITTTR, Chandigarh, India and B.Tech. from NIT,Jalandhar, India. Mr. Mehra has more than 15 years of academic experience. He has authored 35 researchpapers in reputed International Journals and 45 research papers in National and International conferences. Mr. Mehra'sa life member of ISTE.

An Algorithm for Solving Natural Language Query Execution Problems on Relational Databases

Enikuomehin A.O., Okwufulueze D.O.
Dept. of Computer Science,
Lagos State University
Lagos, Nigeria

***Abstract*— There continues to be an increased need for non-experts interaction with databases. This is essential in their quest to make appropriate business decisions. Researchers have, over the years, continued to find a methodology that bridges the gap that exist between information need and users satisfaction. This has been the core in studies related to natural language information retrieval. In this paper, we understudy the existing methodology and develop a model to extend the proposition of (a) Bhardwaj et al where a MAPPER was developed and implemented on student database and (b) Nihalani et al. where an integrated interface was used on relational databases. We present a time saving executable algorithm that satisfies needed conditions required to retrieve results of natural language based queries from relational databases. Results of the experiment shows that the performance index of the algorithm is satisfactory and can be improved upon increasing the metadata table of the relational database. This is a sharp diversion from the keyword based search that has dominated most commercial databases in use today. The implementation was deployed in PHP and the retrieval time has compared favorably with earlier deployed models. We further propose the extension of this work in the areas of inculcating some fuzzy constraints to handle uncertainty and ambiguity which are inherent in human natural language.**

Keywords- Relational Database; Interface; Natural Language; Query; SQL.

I. INTRODUCTION

Research work on developing a flexible Natural Language Interface for Relational Databases has experienced expansion at a very high rate [1]. This has led to continuous research on natural language interfaces and query execution related issues. However, the attention received in this area has not led to significant and commensurate improvement in the existing models for natural language information retrieval essentially in the areas related to development of human useable interfaces. This complexity has been linked with the discreetness required for information extraction from relation databases by the autonomous use of Structured Query Language (SQL). SQL (Structured Query Language) is the formal querying language for relational databases. This is an expert language that is; users need to learn a specific syntax to initiate an appropriate query. In contrast, most business individuals are not experts in this domain and have causes to relate with the relational databases. Obviously, there is a need for this category of users to interact consistently with the content of the relational databases. This paper discusses some of the approaches that had been introduced to enable users query the database using their natural languages rather than SQL. These developed approaches enable database queries to be performed by users with little or no SQL querying abilities. However, some of the systems developed so far are not flexible enough to deal with the complexity associated with human users. Such earlier propositions force the user to adhere to strict grammatical rules when formulating queries. For appropriate usable results to be achieved, queries must be well posed against the relational database. The NLIDB will assist users to reformulate a natural language query into an appropriate SQL. The use of NLIDB has experienced rapid growth and continues to enjoy great support in terms of research and contributions.

If the above holds, one wonders why it is necessary to put some and energy in studying this process with the level of attention received. The answer is simple: the information seeking task becomes more complex and the available number of information object increases. This increment is being experienced by the day with the continuous exponential growth of the internet. This consideration clearly establishes that the existing tools for SQL generation may not be appropriate for some strictly defined domains; we therefore propose an algorithm that is flexible for extension to handle the information growth. In Enikuomehin et al [9], a proposal for handling natural language queries in LANLI was proposed. The resulting implementation performed considerably better than existing commercial interfaces however the time of execution has been a concern to researchers. The formalism involves that non SQL experts could pose a query which runs through a preprocessor. We advance on this proposition to save time and present a direct executable algorithm for natural language retrieval

II. BACKGROUND

Relational Databases (*a collection of data items organized as a set of tables for easy storage, manipulation and retrieval of data*) are becoming ubiquitous as there continues to be an increased need for people - mostly laypeople – to query databases and gain access to information. There is hardly any existing institution today that does not make use of a relational database in managing the massive amount of data the institution deals with. Such cases can be made for Government, Education, Religion, and Business amongst others. These relational databases however can be accessed using formal methods, which require a great deal of learning on the part of the user. This requirement is actually

challenging because, a user who is a novice in the methods used to access a database will find it really difficult to gain access to important information he/she may need at the moment. For example, consider a situation where an expert in database access could not perform his/her duties due to technical incompetence in the formulation of SQL queries. In the early generation of computers, a lot of skills, gotten from a formal and rigorous training in computer usage was required to operate the computer. Subsequent generations dealt with this rather difficult demand of an expert operator, and brought about an era where the less experienced could also operate the computer. To access a database, user must make use of a formal language which the relational database understands. One of such a formal language used to communicate with a database is SQL (Structured Query Language). The use of SQL requires some level of expertise, such expertise are normally acquired after due training. This paper presents a simple and easy-to-use natural language interface to enable less non technical users to have the capability to retrieve information from the relational database.

III. Some Related Earlier Works

Research in Natural Language Interface for Relational Databases began as far back as the 20th century. Since then the study and interest has continued to grow tremendously such that the area has become the most active in Human-Computer Interaction. The first Natural Language Interface for Relational Databases appeared in the 1970s*[2]*, the NLIDB system was called LUNAR*[]*. After the development of the first NLIDB, many were built which were supposed to be an improvement on the apparent flaws of LUNAR. The presentation and acceptance of LUNAR was huge. The reason for such huge success with NLIDBs includes the fact that there are real-world benefits or payoffs that can be derived from this area of study, other fact is that the earlier experimented domain was a single domain where execution of non complex systems are easy and easily adaptable. Same feet were not achieved in the area of using complex databases. [3] we highlight below, the development of some NL interfaces.

A. Lunar (1971)[4]

Man had accomplished the complex task of both having a physical presence on the moon and that of positioning satellites in space that can bring results from observations done on the moon. Information of rock samples brought back from the moon, for example, chemical information were stored in a database, while literature reference on various samples were stored in another database. LUNAR helped provide answers to queries about any of the two information about a rock sample by the use of these databases. LUNAR had linguistic limitations and was able to handle 78% of user-requests.

B. Philiqa [Philips Question Answering Machine](1977)[5]

This system works by having a clear-cut distinction of the syntactic parsing and semantics of the user-defined query. It has three layers of semantic understanding:

a. English Formal Language

b. World Model Language

c. Database Language

Together, these three layers work to answer user-defined queries. Users did not achieve so much acceptance as the earlier developed LUNAR.

C. Ask (1983)[6]

Ask was a complete information management system with an in-built database and the ability to communicate with multiple external databases using several computer applications which are accessible to users through the user's natural language query. Learning is the ability of a system to experience change based on a certain experience with an input such that it can perform an activity better and more efficiently next time. Since ASK had the ability to be taught new concepts by the user during conversation with the user, it can be said that ASK was a learning system.

D. Team (1987)[7]

TEAM was an NLIDB whose developers concerned themselves with portability issues, as they wanted it to be easily implementable on a wide range of systems without compatibility issues. It was designed to be easily configured by database administrators with no knowledge of NLIDB. These feet affected the functionality of TEAM.

E. Precise (2004)

PRECISE introduced the concept of Semantically Tractable Sentences which are sentences whose semantic interpretation is done by the analysis of some dictionaries and semantic constraints.

It was developed by Ana-Maria Popescu, Alexander Yates, David Ko, Oren Etzioni, and Alex Armanasu in 2004 at the University of Washington [8].

When a natural language query is given to PRECISE, it takes the keywords in the sentence of the query, and matches the keywords to corresponding database structures. This, in fact is the major strength of PRECISE. PRECISE does this matching in two stages. The first is to narrow down the possible keywords using the Maximum Flow algorithm which finds a feasible, constraint-satisfying flow through a Flow Network having just a single source and a single sink, such that the flow is maximum; where a flow network is a directed graph in which each edge has a capacity and each edge receives a flow. By using the Maximum Flow algorithm, the maximum number of keywords is obtained, thereby increasing the chance of the natural language sentence to be accurately transformed to a formal SQL query as there will be enough keywords to compare with the PRECISE dictionary. The second stage is to analyse the syntactic structure of the sentence. PRECISE also has its own limitations.

Generally, some major flaws have been common to these interfaces and their ability to handle natural language processing. Users' feedback system has not been thoroughly handled in existing systems. Such systems learn when the user prompts command such as save text on the interface. This is worsened by the fact that, though they are considered as a NLI, their knowledgebase has been a concern in recent times such that can only get results that keyword based. The area of natural language that can be handled by NLIDBs is just

a small subset, and this subset is even difficult to define due to Natural language complexity and the existence of ambiguity.

IV. Overview Of The Proposed System

In Enikuomehin et al[9], an NLI for RDB was developed. In that work, the system developed was named LANLI where a set of operations is defined on a Local Appropriator. The local appropriator allows for both semantic and syntactic tree generation for query execution. The highlight of the proposition is that the matching algorithm would have been generated before the query formulating tree is used. The advantage is in the area of effective retrieval due to accurate tree formation for both the database and the query. An additional feature of the system is the use of a knowledge dictionary like table where the Natural Language presented by users is assumed to have some knowledge interpretation. Same is similar to the work of NIHALIA et al[10] where an interface was designed on plain relational database. The common factor in the above schedule is that they both operate as a query executor that does not require any formal syntactic presentation. In the implementation of the proposed algorithm, the following process is undergone:

- ✓ User's natural language queries are accepted as input to a given natural language interface.
- ✓ Data-transformation of the natural language query into a formal SQL query is performed by an underlying program without the knowledge of the user.
- ✓ The SQL query is then delivered to the relational database.
- ✓ The result of the query produced by the database is accepted and transformed back into expressions in the user's natural language by the underlying program.(this is the reverse operation of 3 define above).
- ✓ This transformed result is then displayed to the user as output.

The system can be integrated into the module of existing commercial systems. The steps outlined above are necessary for an efficient operation of an NLIDB. For experimental purposes, the lecturer- course database of the department of computer science is used a case study for the implementation of the algorithm. The system is a combination of a database and set of tables resident in it. This work introduces the use of corpus in areas other than the strict information retrieval domains. The execution process can be classed into phases and presented as follows:

A. Input To The Natural Language Interface

To use an NLIDB, there must be a point of interaction between the user and the system. This point of interaction must be able to accept data (query in this case) in a form expressed in the natural language of the user, and it must be able to produce output in the same natural language format.

Because it is a point through which users can communicate with the system using their natural language, it is therefore called a Natural Language Interface. It should be noted that for the purpose of this paper, the natural language used is English Language. Thus the Natural Language Interface to be used in this work is one that accepts English Language queries as input.

B. Transformation Of Natural Language Query To Sql

Natural language is the language used for communication by humans. This language is immediately understood intuitively by humans without any further interpretation. However, to carry on conversation with any component of the computer system such as a database, one must make use of some formal language which requires some special kind of rigorous learning process for anyone to have a mastery over it. Expressions in this rather artificial language must conform to some unambiguous syntactic rules for there to be a meaningful conversation between the human and the computer system. Interaction with a database requires the use of a formal language, whose expressions, unlike natural language expressions, contain no ambiguities. Several Database Management Systems DBMS) have their corresponding language used to interact with them. The database used in this project is the Relational Database. To interact with a relational database, the language to be used is Structured Query Language (SQL). Since the natural language interface collects natural language expressions as input, this input has to be converted to a corresponding SQL expression before the database could understand the query of the user. Therefore, there must exist a program or application whose job is to retrieve the natural language input from the Natural Language Interface, and do some transformation works on it to convert it to an equivalent SQL query.

This application should be able to split the natural language query into its constituent tokens, and through comparisons with the contents of the corpus, it should be able to single out keywords in the statement. With the use of these keywords, and the use of a knowledge base (If-Then knowledge base as used in this paper), the user's query should be able to be parsed semantically, enabling the formulation of a corresponding SQL query which will then be passed to the database. The use of a knowledge base implies that the system will be domain dependent, thus it has to be reconfigured for any new database system on which it is implemented. The SQL query resulting from the transformation performed on the natural language query will have to be passed from the application to the database system itself. This transfer is possible if there is an interface between the application and the database system. This interface is usually inbuilt as a class or subroutine in many programming languages. Thus the language used for the application must possess the capability to connect to the database. After processing the SQL query, the RDBMS returns a result, this result set, occurring in less-human-understandable format, should be manipulated, to enable presentation in a natural language format. This is done by the intermediate application program between the interface and the database. In the human-readable format, the results are then ready to be presented to the user.

V. Design And Structure Of The Proposed Nlidb

Five steps are taken in the use an NLIDB and are described below:

A. *User's Natural Language Queries Are Accepted As Input To The Natural Language Interface:*

The interface actually is the first thing a user should encounter. Then the user gets started with the system by entering a query in his/her natural language

B. *A Data-Transformation Of The Natural Language Query Into A Formal Sql Query Is Performed By An Underlying Program:*

In this stage, it will be observed that the natural language query of the user is fed into the underlying application program, which in turn transforms the user's natural language query into an appropriate SQL query. Thus, there exist an interface between the Natural Language Interface and the underlying application program. This interface is responsible for presenting the natural language query from the user to the application. This interface is for the sake of this project called **NL-Application Program Interface (NLAPI)**.

C. *The Sql Query Is Then Delivered To The Relational Database:*

After the transformation of natural language query into Structured Query Language, the application program having first established a connection to the relational database, will now transfers the SQL query to the RDBMS. The connection established between the application program and the database is made possible by the help of another interface called **Application Program-Database Interface (APDI)**. This interface does the presentation of the corresponding SQL query produced by the application program to the RDBMS.

D. *The Result Of The Query Produced By The Database Is Accepted And Transformed Back Into Expressions In The User's Natural Language By The Underlying Program:*

This process is performed by the application program. The application program receives the result of the SQL query, and transforms it back into a form easily understandable by a human user.

E. *This Transformed Result Is Then Displayed To The User As Output:*

The interface here can be viewed as a reverse automated machine that displays the output of the search process. This makes the entire database search a cycle-like process.

VI. An Algorithm For Implemetation

The first thing to be done with a user's query, is to tokenize the words in the user's queries into the words found in the corpus and the requests tables of the database. This tokenization of words is done in such a way that erroneous repetitions are eliminated. The algorithm for the execution is given as:

```
query=the user's query ;
tok=getTheFirstToken(query) ;
 i=0; j=0;
while (tokenStillExists(query)) {
        if(existsInCorpus(tok)){
                toUse[i]=tok;//This array contains
words found in the user's query and also in the
corpus
                i++;
        }
        if(existsInRequests(tok)){
                reqWord[j]=tok;
        term[j]=TColumnInRequest(tok);
                //TColumnInRequest(tok) is
                the value in the t column of
                requests table for the word
                r=tok
                j++;
        }
        tok= nextToken(query);
}//End of while loop
removeDuplicate(toUse);//Removes duplicates
from the user's query
removeDuplicate(reqWord);//Removes
duplicates from the array of non-entity-reference
terms in the array reqWord[].
removeDuplicate(term);//Removes duplicates from
the array of requested data in the array term[].
```

Now that the user's query have been tokenized and separated into different sets. It must be noted that the user's query now tokenized into the array to Use can contain a combination of general and specific words.

The general words in the array to Use is thus stored in the array G and the specific words in toUse are stored in the array S. This results in four different cases for which the execution of the query differ. These cases are:

sizeOf(G)==0 and sizeOf(S)==0

sizeOf(G)==0 and sizeOf(S)!=0

sizeOf(G)!=0 and sizeOf(S)==0

sizeOf(G)!=0 and sizeOf(S)!=0

A knowledge base is created that caters for any one of the above situations; however, a brief discussion is given here to demonstrate what happens in any of the cases.

1. In the case where sizeOf(G)==0 and sizeOf(S)==0, that is, there are no general and specific words in the arrays G and S, this means that the query of the user does not contain any word in the corpus, thus the query is invalid. This message will be shown to the user.

2. In the case where sizeOf(G)==0 and sizeOf(S)!=0, that is, there are no general words but there are specific words in the query of the user, then two cases arise from this:

sizeOf(reqWord)==0

sizeOf(reqWord)!=0

In the case where sizeOf(reqWord)==0, there exists non-entity-reference words in the user's query, this would lead to the production of an SQL query that selects only the data requested by the user from the csc table, else, a general collection of data is selected from csc table for the data item(s) in the set of specific words S.

In fact, for any of the remaining cases:

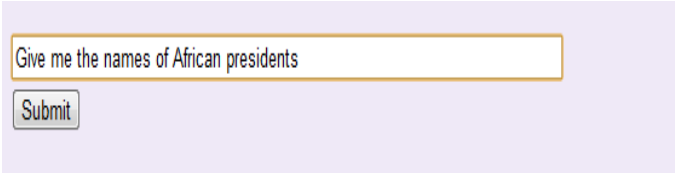
sizeOf(G)!=0 and sizeOf(S)==0

sizeOf(G)!=0 and sizeOf(S)!=0

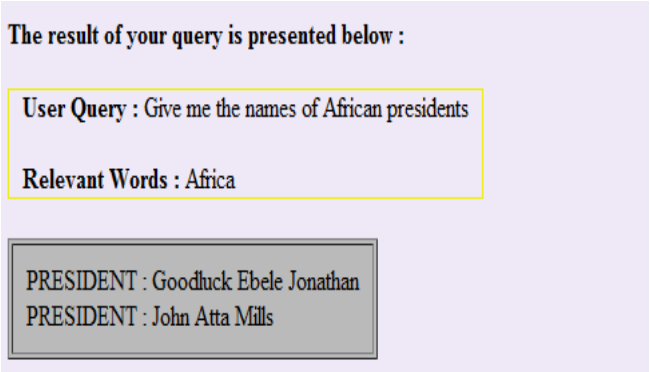
it is tested whether sizeOf(reqWord)==0 or sizeOf(reqWord)!=0, and the codes of the knowledge base found in the intermediate application program does the necessary operations using techniques in both syntactic and semantic parsing to transform the user's query into a corresponding SQL query.

VII. Implementation And Results

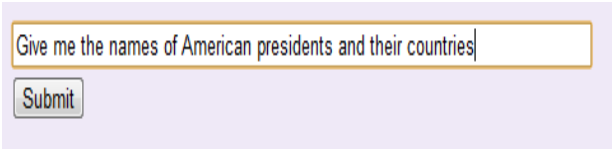
The proposed system is implemented as a web-based application. Thus the languages used include HTML, CSS, JAVASCRIPT, PHP,SQL, while the database used is MySQL as stated earlier. The system answers most of the questions posed to it by the user in natural language. The system enables a user to get information about subject of interest by typing the text in its natural language form. Below is a snapshot of some search carried out to test the performance of the result. a student or lecturer by just typing the latter's phone number or any identifying data for that matter. The snapshots below show some correct inputs and their associated results for the testing on correct inputs.



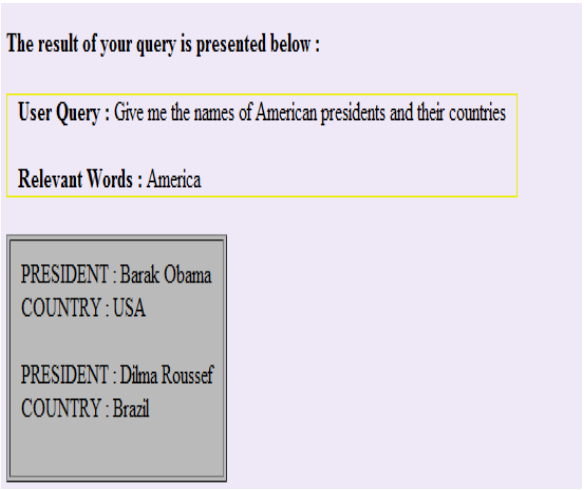
Query 1



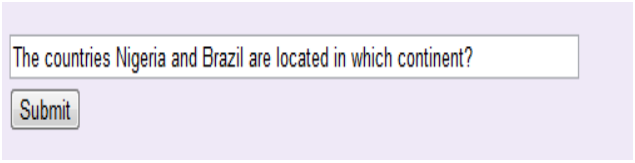
Result 1



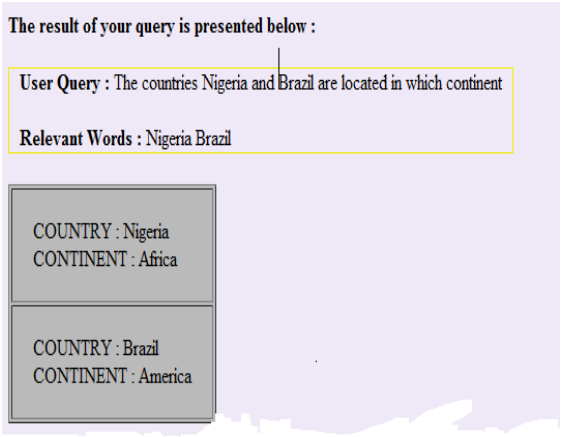
Query 2



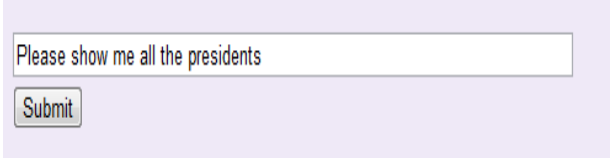
Result 2



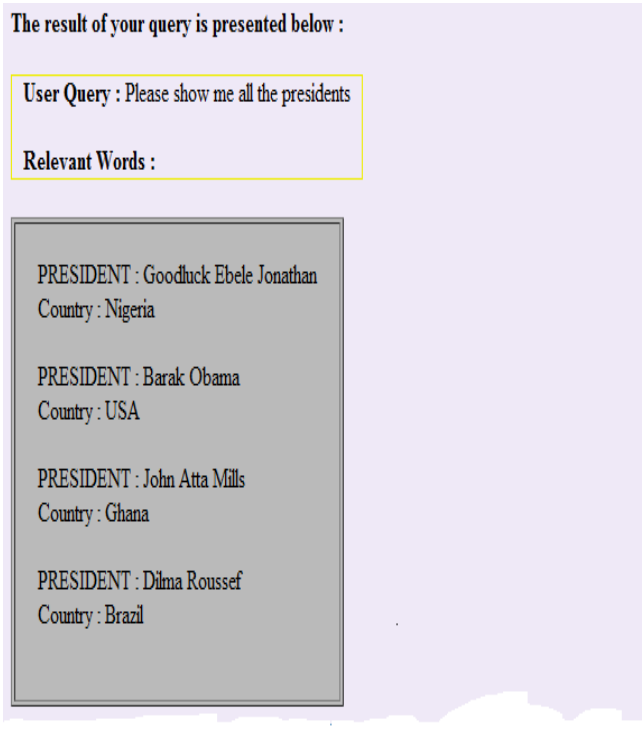
Query 3



Result 3

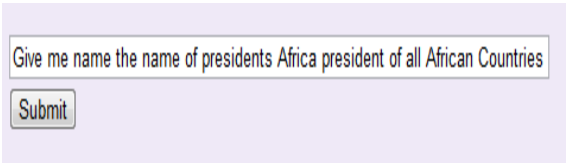


Query4

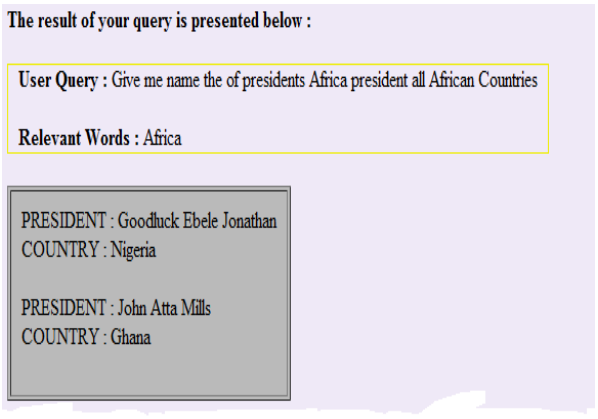


Result 4

The following snapshots below show how cases of grammatically incorrect queries are handled. These errors might arise from the fact that the user has poor grammatical abilities when it comes to the use of English Language [10], or the user forgets that a particular word had been typed once earlier, and then proceeds to type it again.



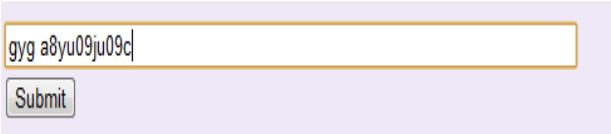
Query 5



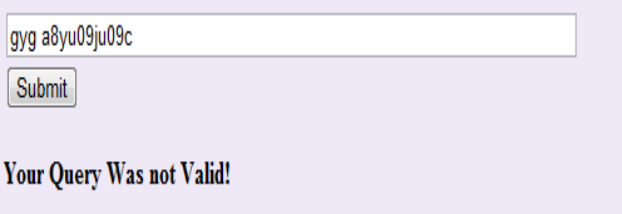
Result 5

In a case where a user just enters an arbitrarily random query, that makes no sense whatsoever in the English Language, the NLIDB should not crash, rather, it should neatly handle this error and show an appropriate message flagging off that error.

The snapshot below shows the result of this idea.



Query 6



Result 6

VIII. Discussions

The flexibility of the Natural Language Interface for Relational Databases is of great importance since it is almost unavoidable for users to make either typographical errors or input out-rightly wrong queries altogether.

A flexible NLIDB should be able to get along somehow with these errors as neatly as possible. This means that there shouldn't be any query whatsoever that could crash the NLIDB.

Flexibility of an NLIDB also makes the computer appear intelligent. This is the main goal of the field of Artificial Intelligence, as a branch of Artificial Intelligence, Natural Language Processing (NLP), and attempts to make human-computer interaction as easy as possible *[4]*. From the experimental results presented above, it is clear that the developed NLIDB is indeed flexible as intended.

It is this flexibility that this project seeks to accomplish and experimentation with random queries have yielded a very high efficient performance rate. The developed NLIDB has its own limitations. These limitations include the following:

- Domain Dependence: The NLIDB is meant to be implemented on a particular Relational Database domain, if it is to be moved to another RDBMS domain, it will have to be reconfigured for that domain. This is one limitation.
- Selective Query Domain: The NLIDB does not answer ALL the questions users may have about different countries of the world. For example, questions on civil issues of different countries will not be answered, as they are beyond the scope of the NLIDB, albeit, such questions can be answered if

words describing such civil issues are included in the corpus.

Despite these limitations, the developed NLIDB have proven to have a high performance rate when it comes to the queries posed to it from its query domain.

The limitations of the developed NLIDB, as stated earlier are as follows:

- Domain Dependent (the goal of most researchers is to design a domain independent NLIDB)[11].
- Limited on Query Domain

However, despite these limitations, the developed NLIDB have proven to have a high performance rate when it comes to the queries posed to it from its query domain, as demonstrated in the previous section on implementation and testing, by experimentations with random selection of queries.

IX. CONCLUSION

Natural language has been successfully to perform a full knowledge based semantically conscious search on relational database. This is the intent of this work. The paper showed how a modelled algorithm can be used to create a user friend non expert search process. The modularity of sql conversion was also shown.

Proposal was implemented on a departmental database however the interest in this work is not the size of the corpus but the time of execution of any unit query. Our proposed model has been able to intelligently process users request in a reasonable human useable format. The implemented result shows that the time is considerable better than earlier propositions and shall thus be upheld.

The research in this area is still ongoing and many interesting additions will be made in the future especially in the area of uncertainty in user information request definition.

REFERENCES

[1] P. Reis, N. Mamede, and J. Matias, "Edite - A Natural Language Interface to Databases: a New Dimension for an Old Approach", Proceeding of the Fourth International Conference on Information and Communication Technology in Tourism, Edinburgh, Scotland, 1997.

[2] M. Minock, "A phrasal approach to natural language access over relational databases", proceedings of the 10th International Conference on Applications of Natural Language to Information Systems, Alicante, Spain, 2005, pp. 333-336

[3] Adam: Student Debt Advisor, Convagent Ltd, Manchester, UK, 2001, Available at: http://www.convagent.com/convagent/adam3.aspx

[4] W. Woods, R. Kaplan, and B. Webber, "The Lunar Sciences Natural Language Information System", Final Report, Technical Report 2378, Bolt Beranek and Newman Inc., 1972

[5] R.J.H., Scha., "Philips Question Answering System PHILIQA1", In SIGART Newsletter, no.61. ACM, New York, (February 1977)

[6] W. Wahlster, H. Marburger, A. Jameson, and S. Busemann. Over-answering Yes-No Questions: Extended Responses in a NL Interface to a Vision System. In: Proc. of the 8th IJCAI, pp. 643–646, Karlsruhe, FRG, 1983

[7] B.J. Grosz, "TEAM: A Transportable Natural-Language Interface System", In Proceedings of the 1st Conference on Applied Natural Language Processing, Santa Monica, California, (1983), pp 39–45

[8] A.M. Popescu, O. E. , and H. Kautz," Towards a Theory of Natural Language Interfaces to Databases " University of Washington Computer Science Seattle, WA 98195, USA.

[9] A. Enikuomehin, J.Sadiku, A new Architecture for NLIDB Using Local Appropriator Engine for SQL Generation, International journal of Advance research in computer science, 2012.

[10] N Nihalani et al ," An Intelligent Interface for Relational Databases", IJSSST, Vol. 11, No. 1, ISSN: 1473-804x online, 1473-8031 print, p30.

[11] R. Ahmad "Efficient Transformation of a Natural Language Query to SQL for Urdu", Proceedings of the Conference on Language & Technology 2009, p53.

Quantifiable Analysis of Energy Efficient Clustering Heuristic

Anita Sethi[1]
[1]Uttrakhand Technical University,
Dehradun, India

J. P. Saini[2]
[2] M.M.M. Engineering College
Gorakhpur, India

Manoj Bisht[3]
[3]WWIL Ltd.
Delhi, India

***Abstract*— One of the important aspects of MANET is the restraint of quantity of available energy in the network nodes that is the most critical factor in the operation of these networks. The tremendous amount of energy using the mobile nodes in wireless communication medium makes Energy Efficiency a fundamental requirement for mobile ad hoc networks. The cluster-based routing protocols are investigated in several research studies which encourage more well-organized usage of resources in controlling large dynamic networks. Clustering can be done for different purposes, such as, routing efficiency, transmission management, backbone management etc. Less flooding, distributed operation, locally repairing of broken routes and shorter sub-optimal routes are the main features of the Clustering protocol. In this paper, we present quantifiable analysis of energy efficient cluster-based routing protocol for uninterrupted stream queries.**

Keywords- Cluster; ClusterHead; Gateway; Associated nodes; Energy Efficiency.

I. INTRODUCTION

A variety of static clustering heuristics permits us to expertly identify group structures. Dynamic nature of MANET signifies the change of a network in the course of discrete time. Clustering technique exhibits a clustered structure based on intra-cluster density versus inter-cluster sparsity of edges and Random location of mobile nodes is generated according to a probabilistic model. Formalizations of this notion lead to measures that quantify the quality of a clustering and to algorithms that actually find clustering. Since, most generally, corresponding optimization problems are hard, heuristic clustering algorithms are used in practice, or other approaches which are not based on an objective function.

A clustering **C(ζ)**of a graph G = (V, E) is a partition ofV, into disjoint, non-empty subsets$\{C_1, C_2, \ldots\ldots, C_k\}$.Each subset is a cluster $C_i, \in \zeta$. *The number of clusters in a clustering with* k = |ζ|. It is convenient to denote the cluster to which u currently belongs by ζ (u)

$$\zeta: \begin{cases} V \to \zeta \\ v \mapsto \mathcal{C} \end{cases}$$

The set of intra-cluster edges of a cluster **C** is defined as

E(C) = {{u, v} ∈ E: u ∈C ⋀v ∈C}

The set of inter-cluster edges between clusters **Ci** and **Cj** is defined as

E (Ci, Cj) = {{u, v} ∈ E: u ∈Ci ⋀v ∈Cj} wherei ≠ j

For a graph **G = (V, E),**an optional weight function**ω**and a coverage clustering ζof G, coverage is defined as

$$\mathcal{C}(G, \zeta) = \sum_{\mathcal{C} \in \zeta} \frac{|E(\mathcal{C})|}{|E|}$$

$$\mathcal{C}_{\omega}(G, \zeta) := \sum_{\mathcal{C} \in \zeta} \frac{\omega(E(\mathcal{C}))}{\omega(E)}$$

For a graph**G = (V, E),**and a clustering**C(ζ)**performance is defined as

$$\mathcal{P}(G, \mathcal{C}) = \frac{|E(\zeta)| + |\bar{E}(\zeta)^c|}{\frac{1}{2} \cdot n \cdot (n-1)}$$

Significance Performance for a graph**G = (V, E),**and a clustering **C(ζ)**is defined as

$$\mathcal{SP}(G, \mathcal{C}) = \mathcal{P}(G, \mathcal{C}) - \mathcal{E}[\mathcal{P}(G, \mathcal{C})]$$

For a graph**G = (V, E),***a* **cut θ = (U,V\U)**is a partition of the node set into two subsets. For a weighted graph**G = (V,E, ω)**, the weight**ω(θ)**of the cut is defined as

$$\omega(\theta) = \sum_{\{u, v\} \in E(U, V \setminus U)} \omega(\{u, v\})$$

The conductance weight of one side of the cut is

$$\alpha(U) = \sum_{\{u, v\} \in E(U, V)} \omega(\{u, v\})$$

For a graph**G = (V, E, ω),**and a clustering**ζ** the conductance for the**cut θ = (C, V\C)**is defined as

$$\phi(\mathcal{C}) = \begin{cases} 1 & \text{if } \mathcal{C} \in \{\emptyset, V\} \\ 0 & \text{if } \mathcal{C} \notin \{\theta, V\} \wedge \omega(\overline{E(\mathcal{C})}) = 0 \\ \frac{\omega(\overline{E(\mathcal{C})})}{\min(\alpha(\mathcal{C}), \alpha(V \setminus \mathcal{C}))} & else \end{cases}$$

For a graph **G = (V, E),** and a clustering **C (ζ)**inter-cluster-conductance is defined as

$$ICC(\zeta) = 1 - \sum_{C_i \in \zeta} \frac{\emptyset(C_i)}{|\zeta|}$$

Dynamic Clustering Heuristic is a technique which, given the previous state of a dynamic graph Gt-1,a sequence of graph events Δ(Gt-1,Gt)and a clustering ζ(Gt-1) of the previous state, returns a clustering ζ(Gt)of the current state.

II. Objectives of Clustering

Aim of the Clustering algorithm is to discover a realistic interconnected set of Clusters covering the entire node population. If the goal is to create an indiscernible global infrastructure where mobile devices can communicate with each other effectively, reliably and wirelessly without utilizing huge amount of energy and no loss of connectivity and data then clustering algorithms is the only way. An efficient Clustering algorithm should be steady to the radio motion, i.e. it should not change the Cluster Configuration too drastically when a few nodes are moving and the topology is slowly changing.

Fig. 1 Clustering Objectives

Otherwise, the Centroids will not control their Clusters efficiently and thus lose their role as local coordinator.

III. Clustering Operation

Clustering process involves three phases,

1) *Initialization*
2) *Setup Phase*
3) *Transmission Phase*

The objective of Cluster Formation is to enforce some kind of structure or hierarchy in the completely disorganized ad hoc network. Originally all nodes wake up in the Ambiguous state in the network initialization phase and use the information obtained from the HELLO messages for Cluster Formation. During Setup phase ClusterHead will be elected according to heuristic which have comprehensive knowledge about group membership and link state information in the Cluster within a bounded time once the topology within a cluster stabilizes. A node regards itself as an associated node for a particular Cluster if it has a bi-directional link to the corresponding ClusterHead in the Setup phase. An associated node may hear from several ClusterHeads and therefore have several host Clusters; its host ClusterHeads are implicitly listed in the HELLO messages it broadcasts.

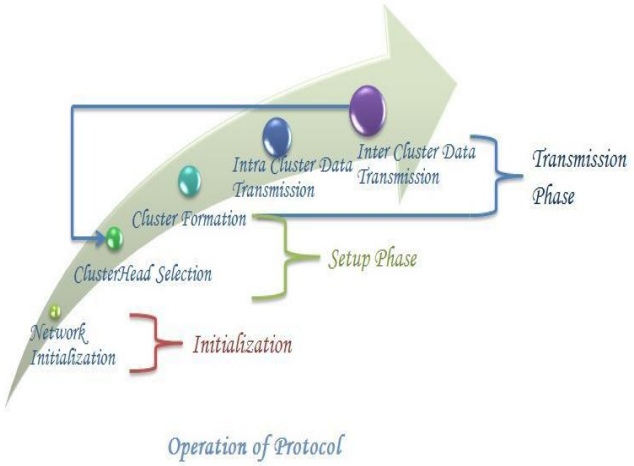


Fig. 2 Clustering Operation

Any node a ClusterHead may use to communicate with an adjacent Cluster is called a Gateway node. As clusters are identified by their respective ClusterHeads, we would like to have the ClusterHeads change as infrequently as possible.

As the nodes unceasingly move in different directions with different speeds, the existing links between the nodes also get changed and hence, the initially formed Cluster cannot be retained for a longer period.

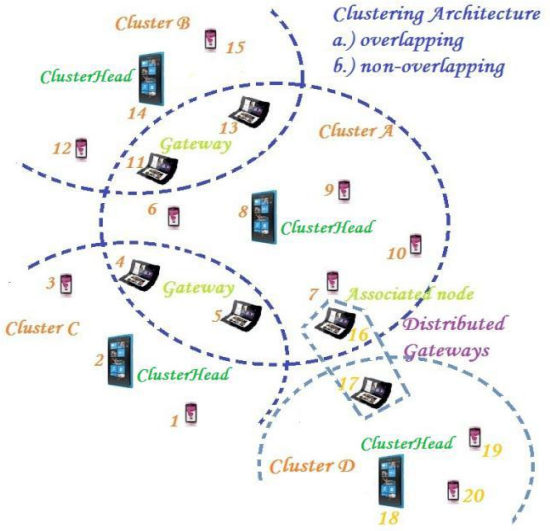


Fig. 3 Clustering Architecture

So, it is necessary to go for the next phase, namely, cluster maintenance phase. Maintenance includes the procedure for modifying the cluster structure based on the movement of a cluster member outside an existing cluster boundary, battery drainage of cluster-heads, link failure, new link establishments, addition of a new node, node failure and so on.

IV. Calculation Of Energy Required For Transmission And Reception Of A Single Packet

Calculation of Time required for Transmission and Reception of a Single Packet		
For Data Packets	Total Packet Size = Size of (preamble + PLCP header + MACK Header + IP Header + Data) Packet Length = 1800 bytes, Bit Rate = 250 kbps	
Total Packet Size	(144+48+28*8+20*8+1800*8) bits	
Transmission Time for Single Packet	0.192ms	preamble and PLCP header
	1.344ms	MACK Header + IP Header + Data
Total Transmission time for Single Packet	1.536ms	
For Ack Packets	Total Packet Size = size of (preamble + PLCP header + ACK) Packet length = 14 bytes Bit Rate = 250 kbps	
Total Packet Size	(144+48+14*8)bits	
Total Transmission time for Single Packet	0.304ms	

A. Calculation of Energy Spent

For the simulation, transmission power used is 1.3$\mathcal{M}$W, and reception power is 0.9$\mathcal{M}$Wand various energy cost components are:

$\mathcal{E}_{T\,pck}$= 1.3 * 1.536 * 10^{-3} =1.9968$\mathcal{M}$W

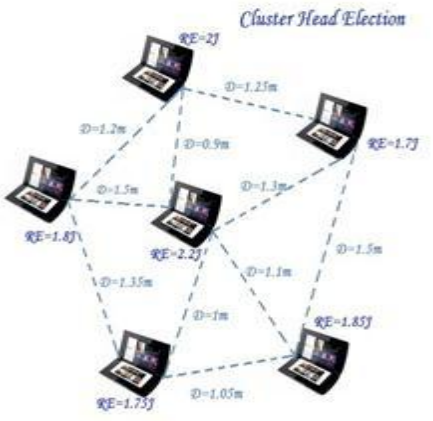


Fig. 4 ClusterHead Election

$\mathcal{E}_{Rpck}$=0.9 * 1.536 * 10^{-3}= 1.3824$\mathcal{M}$W

$\mathcal{E}_{T\,ack}$= 1.3 * 0.304 * 10^{-3} = 0.3952$\mathcal{M}$W

$\mathcal{E}_{Rack}$= 0.9 * 0.304 * 10^{-3} = 0.2736$\mathcal{M}$W

V. Observations

A. Obtaining Ideal number of Clusters

We can obtain ideal no of clusters from the equation

$$k = \sqrt{\frac{n}{2\pi}}\sqrt{\frac{\epsilon_s}{\epsilon_l d^4 - (2m-1)E_e - mE_{BF}}}M$$

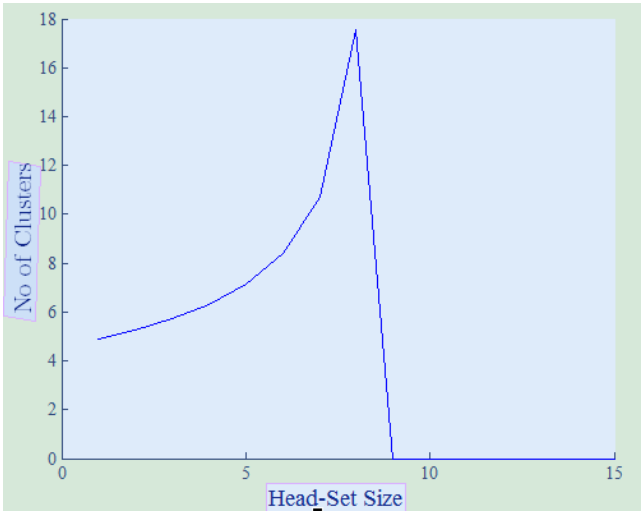


Fig. 5

A simulation environment is created where n = 1200 and d = 160m, n are the nodes and d is the distance between source and destination nodes. The graph drawn with this simulation shows the variation in ideal no of clusters w.r.t the Cluster size. A variation of Cluster size between one and eight is done.

The graph shows that the Cluster size cannot be more than 8 and further we may also derive the maximum no of clusters from the graphical analysis.

In figure 6 a graph is plotted to explain energy consumption wrt number of clusters. We see as the number of clusters is increased the consumption of energy gets considerably reduced. The rate of energy consumption is lowest when more cluster members are there in a cluster. It is clearly visible that when cluster size is 4 the energy cosumption is lowest as compared when cluster size is 1, 2, 3 or4.

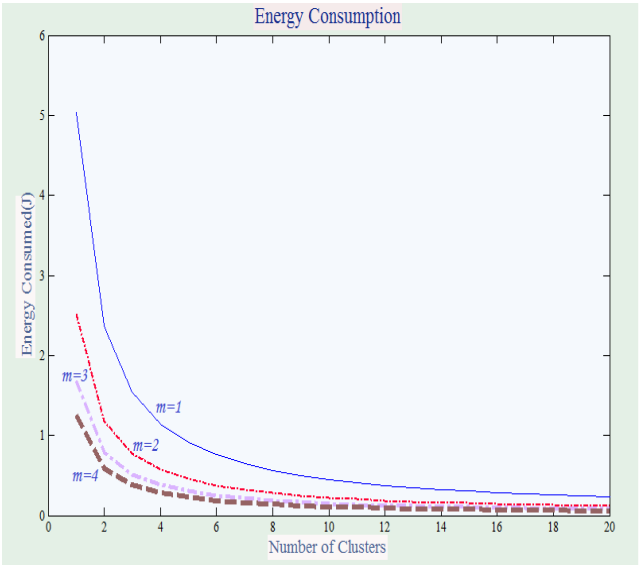


Fig. 6

B. Consumption of Energy

This portion compares the energy consumed for varying head sizes along with definite number of frames. An illustration of variation in energy consumed per node w.r.t. the number of clusters and network diameter is done in figure 7. Numbers of clusters are represented on X-axis while Y-axis represents the energy consumed in a single round. The below mentioned equation is used to derive energy consumed for a single node:

$$E_{start} = E_{CH/node} + \left(\frac{n}{km} - 1\right) E_{non-CH/node}$$

The number of frames transmitted in a single round is 20. The graph clearly depicts that when clusters are increased, energy consumption is reduced. When a simulated network of 1200 nodes is created, graph shows the ideal range clusters lies between 20 and 60. The energy consumption increases when the number of clusters is increased. If the number of clusters is below the ideal range, for e.g. 10, the data collecting sensor nodes have to communicate to the far ClusterHeads and resultantly use more energy in transmission. Similarly if the number of Clusters is more than the ideal value there will be more transmission to the destination. From the graph it is clearly visible that the energy consumption is lower for higher Cluster-size. We can observe that the energy consumed is almost 3 times less when Cluster-size is 1.

$$E_{start} = E_{CH/node} + \left(\frac{n}{km} - 1\right) E_{non-CH/node}$$

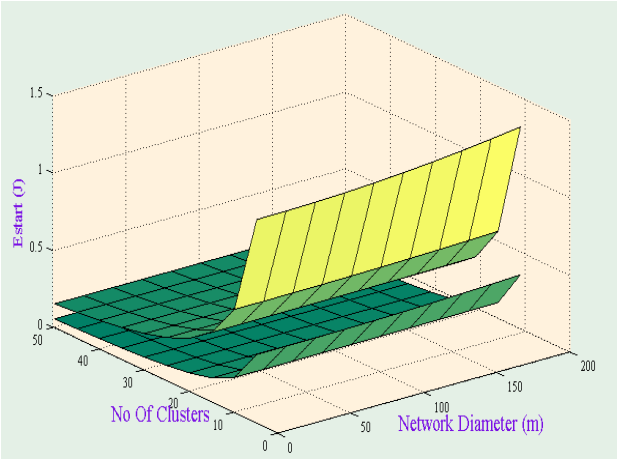


Fig. 7

We can dedicate the main reason to reduction in energy consumption mainly to excess nodes in the Cluster. During the transmission of each frame there are nk-m Non-ClusterHead transmission and one ClusterHead transmission. Further m-1 nodes are in sleeps and do not transmit.

As compared to LEACH there are lesser elections and re-affiliations in the routing model; the number of elections reduced from n/k to n/km. An illustration of per round energy consumed w.r.t cluster size and network diameter is done in figure 8. Network diameter, Head set size and the energy consumed in one round are respectively shown by x, y and z-axis. Energy consumed per round by

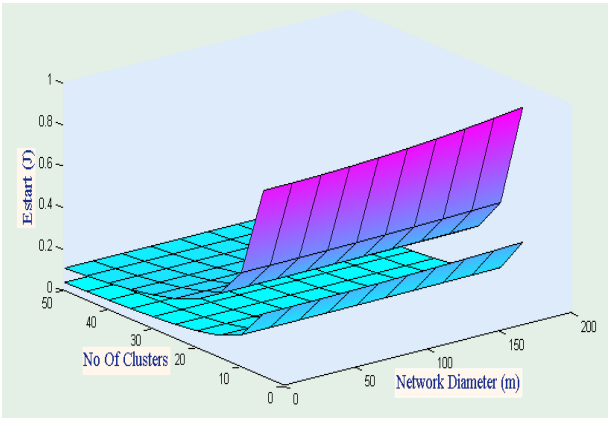


Fig. 8

In one iteration the number of data frames are N_f=10,000 and the number of clusters K=50. The graph clearly depicts that energy consumption is reduced when the head set size is increased. It is very much clear that using a cluster of associated nodes is more beneficial as compared to using single ClusterHead. There is more pragmatic approach in reducing energy consumption when we use this protocol. As we add more nodes to LEACH, they all are treated in differentially and all these nodes will be used for collecting the sensor data. The number of sensor nodes for data collection remains unaltered and the number of control and management nodes can be tuned.

C. Data Frames and Elapsed time in iteration

An estimation of average time for one iteration in each round is done such that every node becomes a member of ClusterHead. Further frames transmitted in each iteration are also evaluated. The variation in time to complete one iteration w.r.t the cluster diameter and ClusterHead size completion time of one iteration are being represented by X, Y, Z axis respectively. The initial energy $\mathcal{E}_{start}$ is fixed for all the cases. When the ClusterHead size is 50% of the cluster size, then the initial energy can be used for longest time interval. When the ClusterHead size is less than 50% of cluster size, then there is less transmission in each iteration.

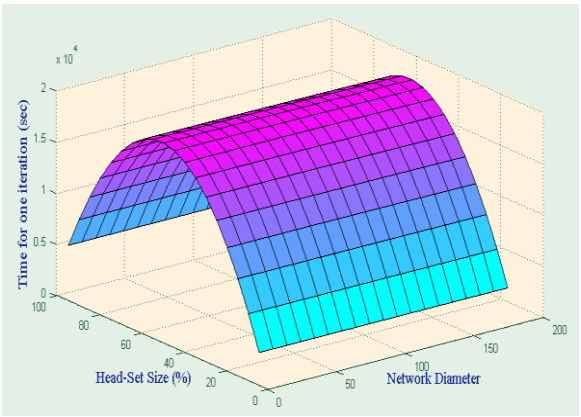


Fig. 9

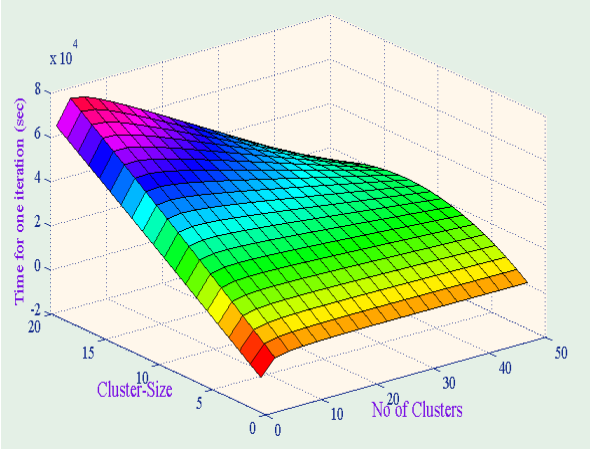


Fig. 10

VI. CONCLUSION

Clustering Heuristic is a powerful and scalable heuristic for ad hoc network. The results of our quantifiable analysis of the energy efficient clustering heuristic indicates that the energy depletion can be analytically decreased by including more nodes in a head-set. For the same number of data collecting, the number of control and management nodes can be adjusted according to the network environment.

In future work, the variation in the head-set size for different network conditions will be investigated.

This work will be prolonged to integrate non-uniform cluster distributions. We are developing the simulation model to authenticate and confirm our quantifiable analysis.

REFERENCES

[1] S. Bandyopadhyay and E. J. Coyle. An energy efficient hierarchical clustering algorithm for wireless sensor networks. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, 2003.

[2] Cerpa and D. Estrin. ASCENT: Adaptive self-configuring sensor networks topologies. *IEEE Transactions on Mobile Computing (TMC) Special Issue on Mission-Oriented Sensor Networks*, 3(3), July-September 2004.

[3] M. Jiang, J. Li, and Y. C. Tay, "Cluster Based Routing Protocol(CBRP) (INTERNET-DRAFT draft-ietf-manet-cbrp-spec-01.txt)," in *National University of Singapore*, I. E. T. F. (IETF), Ed., 1999, pp. 1-27.

[4] S. A. Hosseini-Seno, B. Pahlevanzadeh, T. C. Wan, R. Budiarto, and S. Ramadass, "Routing Layer Service Advertisement Approach for MANETs," in *International Conference on Future Networks (ICFN 2009)* Bangkok- Thiland: IEEE, 2009, pp. 249-254.

[5] Christian Staudt "Experimental Evaluation of Dynaic Graph Clustering Algorithms"

[6] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan. Energy efficient communication protocol for wireless microsensor networks. In *Proceedings of the Hawaii International Conference on System Sciences*, January 2000.

A Newer User Authentication, File encryption and Distributed Server Based Cloud Computing security architecture

Kawser Wazed Nafi[1,2], Tonny Shekha Kar[2], Sayed Anisul Hoque[3], Dr. M. M. A Hashem[4]
[1]Lecturer, Stamford University, Bangladesh
[2]Khulna University of Engineering and Technology
[3]Chittagong University of Engineering and Technology
[4]Professor, Khulna University of Engineering and Technology

***Abstract*— The cloud computing platform gives people the opportunity for sharing resources, services and information among the people of the whole world. In private cloud system, information is shared among the persons who are in that cloud. For this, security or personal information hiding process hampers. In this paper we have proposed new security architecture for cloud computing platform. This ensures secure communication system and hiding information from others. AES based file encryption system and asynchronous key system for exchanging information or data is included in this model. This structure can be easily applied with main cloud computing features, e.g. PaaS, SaaS and IaaS. This model also includes onetime password system for user authentication process. Our work mainly deals with the security system of the whole cloud computing platform.**

Keywords- Cloud Computing; Security architecture; AES; RSA; onetime password; MD5 Hashing; Hardwire database encryption.

I. INTRODUCTION

At the present world of networking system, Cloud computing [1] is one the most important and developing concept for both the developers and the users. Persons who are interrelated with the networking environment, cloud computing is a preferable platform for them. Therefore in recent days providing security has become a major challenging issue in cloud computing.

In the cloud environment, resources are shared among all of the servers, users and individuals. As a result files or data stored in the cloud become open to all. Therefore, data or files of an individual can be handled by all other users of the cloud. [2, 3] Thus the data or files become more vulnerable to attack. As a result it is very easy for an intruder to access, misuse and destroy the original form of data. An intruder can also interrupt the communication. Besides, cloud service providers provide different types of applications which are of very critical nature. Hence, it is extremely essential for the cloud to be secure [4]. Another problem with the cloud system is that an individual may not have control over the place where the data needed to be stored. A cloud user has to use the resource allocation and scheduling, provided by the cloud service provider. Thus, it is also necessary to protect the data or files in the midst of unsecured processing. In order to solve this problem we need to apply security in cloud computing platforms. In our proposed security model we have tried to take into account the various security breaches as much as possible.

At present, in the area of cloud computing different security models and algorithms are applied. But, these models have failed to solve all most all the security threats. [5, 6, 7] Moreover for E-commerce [8] and different types of online business, we need to imply high capacity security models in cloud computing fields. Security models that are developed and currently used in the cloud computing environments are mainly used for providing security for a file and not for the communication system [9]. Moreover present security models are sometimes uses secured channel for communication [10]. But, this is not cost effective process. Again, it is rare to find a combined work of main server security, transaction between them and so on. Some models attempt on discussing about all of these, but are completely dependent on user approach. The models usually fail to use machine intelligence for generating key and newer proposed model. Some models have proposed about hardware encryption system for secured communication system [11]. The idea is usually straightforward, but the implementation is relatively difficult. Besides, hardware encryption is helpful only for the database system, not for other security issues. Authenticated user detection technique is currently very important thing. But, this technique is rarely discussed in the recently used models for ensuring security in cloud computing.

In this paper we have proposed new security architecture for cloud computing platform. In this model high ranked security algorithms are used for giving secured communication process. Here files are encrypted with AES algorithm in which keys are generated randomly by the system. In our proposed model distributive server concept is used, thus ensuring higher security. This model also helps to solve main security issues like malicious intruders, hacking, etc. in cloud computing platform. The RSA algorithm is used for secured communication between the users and the servers.

This paper is formatted in the following way: - section II describes related work of this paper work, section III describes proposed architecture and its working steps, section IV describes the experimental environment, results in different

aspects and advantages of the proposed model, and section V describes the future aspects related to this paper work.

II. Related Work

Numerous research on security in cloud computing has already been proposed and done in recent times. Identification based cloud computing security model have been worked out by different researchers [12]. But only identifying the actual user does not all the time prevent data hacking or data intruding in the database of cloud environment. Yao's Garbled Circuit is used for secure data saving in cloud servers [13, 14]. It is also an identification based work. The flaw in this system is that it does not ensure security in whole cloud computing platform. Research related to ensuring security in whole cloud computing environments was already worked out in different structures and shaped. AES based file encryption system is used in some of these models [15, 16]. But these models keep both the encryption key and encrypted file in one database server. Only one successful malicious attack in the server may open the whole information files to the hacker, which is not desirable. Some other models and secured architectures are proposed for ensuring security in cloud computing environment [17, 18]. Although these models ensures secured communication between users and servers, but they do not encrypt the loaded information. For best security ensuring process, the uploaded information needs to be encrypted so that none can know about the information and its location. Recently some other secured models for cloud computing environment are also being researched [19, 20]. But, these models also fail to ensure all criteria of cloud computing security issues [21].

III. Proposed Model

In our proposed model we have worked with the following security algorithms:-

- RSA algorithm for secured communication [22, 23]
- AES for Secured file encryption [24, 25, 26]
- MD5 hashing for cover the tables from user [27]
- One time password for authentication [28, 29].

At present ensuring security in cloud computing platform has become one of the most significant concerns for the researchers. We have undertaken these problems in our research, to provide some solution correlated with security. We have proposed the following security model for cloud computing data storage shown in Figure 1.

In this model, all the users irrespective of new or existing member, needs to pass through a secured channel which is connected to the main system computer. System server computer has relation with other data storage system. The data storage system can be servers or only storage devices. Here, each of the data storage devices can be thought as one or more servers in number. This means, there are no dedicated servers in cloud computing, rather all are independent servers and can be scaled as necessary.

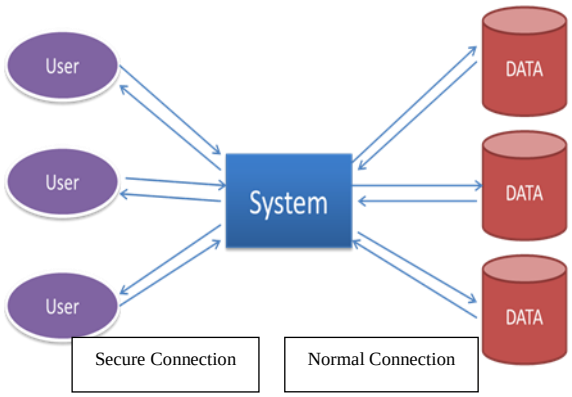


Figure 1. Proposed Security Model

In the proposed model RSA encryption algorithm is used for making the communication safe. Usually the users' requests are encrypted while sending to the cloud service provider system. RSA algorithm using the system's public key is used for the encryption. Whenever the user requests for a file the system sends it by encrypting it via RSA encryption algorithm using the user's public key. Same process is also applied about the user password requests, while logging in the system later. After receiving an encrypted file from the system the user's browser will decrypt it with RSA algorithm using the user's private key. Similarly when the system receives an encrypted file from the user it will immediately decrypt it using its private key. As a result the communication becomes secured between the user and the system.

In the proposed security model one time password has been used for authenticating the user. The password is used to keep the user account secure and secret from the unauthorized user. But the user defined password can be compromised. To overcome this difficulty one time password is used in the proposed security model. Thus whenever a user login in the system, he/she will be provided with a new password for using it in the next login. This is usually provided by the system itself. This password will be generated randomly. Each time a new password is created for a user, the previous password for that user will be erased from the system. New password will be updated for that particular user. A single password will be used for login only once. The password will be sent to the users authorized mail account. Therefore at a same time a check to determine the validity of the user is also performed. As a result only authorized user with a valid mail account will be able to connect to the cloud system. By this system, existence of unauthorized user or a user with an invalid mail account will be pointed out. The newly generated password is restored in the system after md5 hashing. The main purpose of MD5 hashing is that this method is a one way system and unbreakable. Therefore it will be difficult for an unauthorized or unknown party for retrieving the password for a selected user even if gained access to the system database.

After connecting with the system a user can upload or download the file(s). For the first time when connected with the system the user can only upload file(s). After that users can both upload and download their files. When a file is uploaded by an user the system server encrypts the file using AES encryption algorithm. In the proposed security model 128 bit key is used for AES encryption. 192 bit or 256 bit can also be used for this purpose. Here the 128 bit key is generated randomly by the system server. A single key is used only once. That particular key is used for encrypting and decrypting a file of a user for that instance. This key is not further used in any instance later. The key is kept in the database table of the system server along with the user account name. Before inserting the user account name it is also hashed using md5 hashing. This insures that unauthorized person cannot retrieve the key to decrypt a particular file for a particular user by simply gaining access and observing the database table of the system server. As a result the key for a particular file becomes hidden and safe. Again when the encrypted file is uploaded for storing to the storage server, the path of the encrypted file along with the user account is kept and maintained in the database table on the storage server. Here user name is used for synchronization between the database tables of main system server and the storage server. The encrypted files on the storage server are inserted not serially. We have developed a hash table for determining where to insert a file into the database table. The algorithm for generating the hash table is described later in this section.

Login into the main system is compulsory when a user wants to download a previously stored file. When the user selects a file to download, the system automatically retrieves the key for the requested file from the main system server. The system matches user account name saved in its database table with that saved in the storage server after hashing it using md5 hashing. The path of the encrypted file from the storage server is found by using the user account name and the hash table input for the requested file.

In this model, the encryption key for a particular file of a particular user is only known to the main system server. The path of the encrypted file is only known to the storage server which is only known to the main server. For this, the key as well as the encrypted file is hidden from the unauthorized persons. In this communication system when a file is sent from the main system server to the storage server it is already in its fully encrypted form. That's why there is no need to provide security in this communication channel. At last, we propose hardware encryption for making the databases fully secured from the attackers and other unauthorized persons.

Figure 2 is the Pictorial representation of the proposed cloud security architecture. Here, single user and server represent n users and n servers.

An algorithm is developed, which is used for inserting the file in the main server (System), and in the database table where the encrypted file is kept. This is saturated from the system server for the cloud computing platform. In the system server, the file is inserted by maintaining the sequence. In file saving server, the file is inserted in a random order which becomes the output of the algorithm. The relations between the system server table and database server tables can be thought as disjoint sets. The pseudo code of the algorithm used is described in table I.

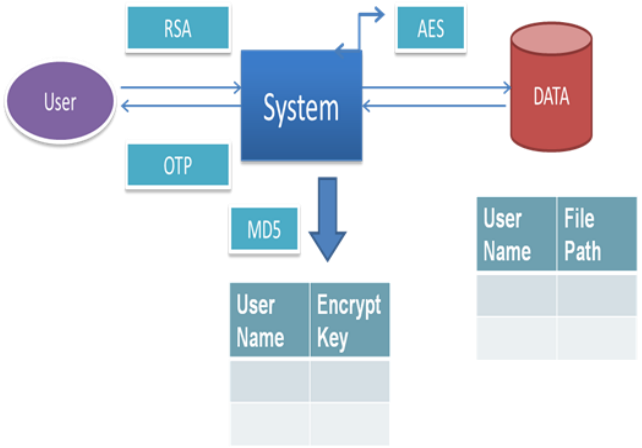


Figure 2. Proposed Security Model/ Structure

TABLE I New Algorithm for uploading file in the Proposed Cloud Architecture:-

The algorithm for generating the hash table which is used for inserting a file in the database table of the storage server is described below:

Step 1: - Select a seed S for generating the hash table which is equal to the block size of the table. Block size means with how many positions of files will be taken from a series of execution

Step 2: - Compute the position where to insert a file.

Position = $N2 \ mod \ S$.

Where N represents the no. of file and S represents the seed value.

Step 3:- a) if Position is empty, then insert the file in that Position.

b) else, increment the Position and set Offset. Repeat step 3.

A sample hash table with seed S = 100 is shown in table II:

TABLE II Synchronization of files in two servers

File No In System Server	Position Of File In Database Server	Offset
1	1	0
2	4	0
:	:	:
:	:	:
5	25	0
:	:	:
:	:	:
15	26	1
:	:	:
:	:	:

IV. EXPERIMENTAL RESULTS

In the lab we have worked with about 100 users and also with their files for studying and prove the efficiency of the proposed model. We have tried to find out different execution results which helped us to demonstrate our model with better result. Different conditions and positions were observed during the working and execution time of this proposed model.

A. Lab Setup

- Platform: Visual Studio 2010 (asp.net)
- Processor: Core 2 Duo (2.93 GHz),
- RAM: 2 GB

In this environment, the whole model took average of 5 seconds for executing all the steps. This hardware configuration takes highest 2 seconds to encrypt about a 10 KB file. This model is fast enough and can be applied to current cloud computing environments.

B. Case Studies

Working with the model in Lab at different times and with different user and their individual files, which are different from each other in size, contents, extension, etc. take different times for executing the overall model. Depending on the file size, program execution time varies from person to person. Among the 100 users result, 10 of them are shown in table III and table IV.

TABLE III Execution time for Uploading File of 10 People

Person No	File Size	Time Required for file Upload (Full Process)	Person No	File Size	Time Required for file Upload (Full Process)
1	1 KB	3 sec	6	17 KB	10 sec
2	4 KB	5 sec	7	15 KB	10 sec
3	14 KB	9 sec	8	5 KB	5 sec
4	7 KB	6.5 sec	9	2 KB	3 sec
5	9 KB	8	10	8 KB	8 sec

TABLE IV Execution time for Downloading File of 10 People

Person No	File Size	Time Required for file Upload (Full Process)	Person No	File Size	Time Required for file Upload (Full Process)
1	1 KB	3.5 sec	6	17 KB	11 sec
2	4 KB	5.5 sec	7	15 KB	11 sec
3	14 KB	10 sec	8	5 KB	5.5 sec
4	7 KB	7 sec	9	2 KB	3.5 sec
5	9 KB	9	10	8 KB	9 sec

From table III and table IV we can see that the proposed model takes quite same time for execution like other present models. But it ensures higher security. Information is stored in main server about the databases where the encrypted files are kept. Thus, database encryption [30, 31] only in main server is enough so that no information is leaked. This makes the model cost effective and less time required for execution of the whole process. Secured information exchanging between the users and system gives protection of hiding information from the unauthorized users and intruders. Comparative analysis of the proposed model is shown in table V.

TABLE V Advantages of the Proposed Model

Points for discussion	**Identification Based Model**	**File encryption based Model**	**Secured channel using model**	**Proposed Model**
Ways of ensuring security	Only identify the authorized person, so hacker can get access on database	Key and file both remains in one server. So, getting access on one server helps to get all information	Intruder cant access the data, but uploaded file is not secured	Ensures security in data exchanging process. Only getting control over full system can leak information
Points for discussion	**Identification Based Model**	**File encryption based Model**	**Secured channel using model**	**Proposed Model**
Information leakage probability	Medium	Medium	Medium	Low
Complexity	Low	Medium	Low	Medium
Cost of establishing and maintaining	Low	Medium	High	Medium
Ensuring User Authentication	Main theme	If key is chosen by user, then slightly authenticate users	Probably not maintained	One time password system is used for user authentication
Execution time	Small	Medium	Small	Medium
Security Breaking probability	Medium	Medium	Medium	Probably Low than others

From the above comparative analysis, we can see that the proposed model works smoothly like others and ensures higher security than other present running models in a cloud computing environment.

V. CONCLUSION

In this paper we have proposed a newer security structure for cloud computing environment which includes AES file encryption system, RSA system for secure communication, Onetime password to authenticate users and MD5 hashing for hiding information. This model ensures security for whole cloud computing structure.

Here, execution time is not subsequently high because implementation of each algorithm is done in different servers. In our proposed system, an intruder cannot easily get information and upload the files because he needs to take control over all the servers, which is quite difficult. The model, though it is developed in a cloud environment, individual servers' operation has got priority here. So, decision

taking is easy for each server, like authenticate user, give access to a file etc.

In our proposed model we have used RSA encryption system which is deterministic. For this reason, it becomes fragile in long run process. But the other algorithms make the model highly secured. In future we want to work with ensuring secure communication system between users and system, user to user. We also want to work with encryption algorithms to find out more light and secure encryption system for secured file information preserving system.

ACKNOWLEDGMENT

The Authors are willing to express their profound gratitude and heartiest thanks to all the researchers in the field of cloud computing architecture's security, specially to the developers of security algorithms, who have made their research work easy to accomplish.

REFERENCES

[1] Yashpal Kadam, "Security Issues in Cloud Computing A Transparent View", International Journal of Computer Science Emerging Technology, Vol-2 No 5 October, 2011 , 316-322

[2] Rohit Bhadauria, Rituparna Chaki, Nabendu Chaki, Sugata Sanyal, "A Survey on Security Issues in Cloud Computing", 2011

[3] Mladen A. Vouk, "Cloud Computing – Issues, Research and Implementations", Journal of Computing and Information Technology - CIT 16, 2008, 4, 235–246

[4] Ye Hu, Johnny Wong, Gabriel Iszlai, Marin Litoiu, "Resource Provisioning for Cloud Computing", IBM Canada Ltd., 2009

[5] Daniele Catteddu, Giles Hogben, "Cloud Computing:- Benefits, risks and recommendations for information security", November, 2009

[6] "Cloud Computing: Silver Lining or Storm Ahead?", Volume 13 Number 2, Spring 2010

[7] NGONGANG GUY MOLLET, "CLOUD COMPUTING SECURITY" , Thesis Paper, April 11, 2011

[8] Gunasekar Kumar, Anirudh Chelikani, "Analysis of security issues in cloud based e-learning", Master's thesis, 2011

[9] Jiyi Wu, Qianli Shen, Tong Wang, Ji Zhu, Jianlin Zhang "Recent Advances in Cloud Security", JOURNAL OF COMPUTERS, VOL. 6, NO. 10, OCTOBER 2011

[10] Ahmad-Reza Sadeghi, Thomas Schneider, and Marcel Winandy, "Token - Based Cloud Computing Secure Outsourcing of Data and Arbitrary Computations with Lower Latency", TRUST 2010, LNCS6101, pp . 417–429, 2010.

[11] Trusted Computing Group, "Solving the Data Security Dilemma with Self-Encrypting Drives", May 2010

[12] Hongwei Li, Yuanshun Dai, Ling Tian and Haomiao Yang, "Identity-Based Authentication for Cloud Computing", CloudCom 2009, LNCS 5931, pp. 157–166, 2009

[13] Sven Bugiel, Stefan Nurnberger, Ahmad-Reza Sadeghi, Thomas Schneider, "Twin Clouds: Secure Cloud Computing with Low Latency", CASED, Germany, 2011

[14] Sven Bugiel, Stefan Nurnberger, Ahmad-Reza Sadeghi, Thomas Schneider, "Twin Clouds: Secure Cloud Computing with Low Latency"- Extended Abstract, CASED, Germany, 2011

[15] Luis M. Vaquero, Luis Rodero-Merino, Daniel Morán, "Locking the sky: a survey on IaaS cloud security", Computing (2011) 91:93–118

[16] Yang Tang, Patrick P. C. Lee, John C. S. Lui, and Radia Perlman, "FADE: Secure Overlay Cloud Storage with File Assured Deletion", 2010

[17] Thuy D. Nguyen, Mark A. Gondree, David J. Shifflett, Jean Khosalim, Timothy E. Levin, Cynthia E. Irvine, "A Cloud-Oriented Cross-Domain Security Architecture", The 2010 Military Communications Conference, U.S. Govt.

[18] Cong Wang, Qian Wang, and Kui Ren, Wenjing Lou, "Ensuring Data Storage Security in Cloud Computing", US National Science Foundation under grant CNS-0831963, CNS-0626601, CNS-0716306, and CNS-0831628, 2009

[19] Vaibhav Khadilkar, Anuj Gupta, Murat Kantarcioglu, Latifur Khan, Bhavani Thuraisingham, "Secure Data Storage and Retrieval in the Cloud", University of Texas, 2011

[20] John Harauz, Lori M. Kaufman, Bruce Potter, "data Security in the World of Cloud Computing", The IEEE Computer SOCIETIES, August, 2009

[21] Kevin Hamlen, Murat Kantarcioglu, Latifur Khan, Bhavani Thuraisingham, "Security Issues for cloud computing", International Journal of Information Security and Privacy, 4(2), 39-51, April-June 2010

[22] R.L. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems", Laboratory for Computer Science, Massachusetts Institute of Technology, Cam-bridge, November, 1977

[23] Burt Kaliski, The Mathematics of the RSA Public-Key Cryptosystem, RSA Laboratories

[24] Joan Daemen, Vincent Rijmen, "AES Proposal: Rijndael", 1999

[25] Joan Daemen, Vincent Rijmen, "Announcing the ADVANCED ENCRYPTION STANDARD (AES)", Federal Information Processing Standards Publication 197, November 26, 2001

[26] Joshua Holden, Mohammad Musa, Edward Schaefer, and Stephen Wedig, "A Simplified AES Algorithm", January 2010

[27] Ronald Rivest, "MD5 Message-Digest Algorithm", rfc 1321, April 1992

[28] Neil M.Haller, "THE S/KEY ONE-TIME PASSWORD SYSTEM", 1993

[29] Neil Haller, "A One-Time Password System", October 23, 1995

[30] "Securing Data at Rest: Developing a Database Encryption Strategy"- A White Paper for Developers, e-Business Managers and IT

[31] Ulf T. Mattsson, "Database Encryption - How to Balance Security with Performance", 2004

AUTHORS PROFILE

Kawser Wazed Nafi passed from Computer Science and Engineering department of Khulna University of Engineering and Technology in June 2012. He then started his career in Samsung Bangladesh R & D centre. Now he is working as Lecturer in Stamford University, Bangladesh in Computer Science and Engineering department. He has been working in Cloud Computing field for more than about one year. He has already published Journal and conference papers on Cloud Computing field in different knowable journals like IEEE, IJCA, IJCOT. His research interest is Cloud Computing, Ubiquitous Computing, Adhoc networks, Artificial Intelligence, Pattern Recognition and So on.

Tonny Shekha Kar passed from Computer Science and Engineering department of Khulna University of Engineering and Technology in June 2012. She has been working in Cloud Computing and Distributed Computing field for more than about one year. She has already published papers on Cloud Computing field, Patteren Recognition, etc in different knowable journals like IEEE, IJCA, IJCOT. His research interest is Cloud Computing, ubiquitous Computing, Adhoc networks, Artificial Intelligence, Machine Learcning, Neural Network, Pattern Recognition and So on.

Sayed Anisul Hoque passed from Computer Science and Engineering department of Chittagong University of Engineering and Technology in March, 2011. He then started his career in Samsung Bangladesh R & D centre as software engineer. His research interest is Cloud Computing, Ubiqutous Computing, wireless network, operating system, android platform and So on.

M. M. A. Hashem received the Bachelor's degree in Electrical & Electronic Engineering from Khulna University of Engineering & Technology (KUET), Bangladesh in 1988. He acquired his Master's Degree in Computer Science from Asian Institute of Technology (AIT), Bangkok, Thailand in 1993 and PhD degree in Artificial Intelligence Systems from the Saga University, Japan in 1999. He is a Professor in the Department of Computer Science and Engineering, Khulna University of Technology (KUET), Bangladesh. His research interest includes Soft Computing, Intelligent Networking, Wireless

Networking, Distributed Evolutionary Computing etc. He has published more than 50 referred articles in international Journals/Conferences. He is a life fellow of IEB and a member of IEEE. He is a coauthor of a book titled "Evolutionary Computations: New Algorithms and their Applications to Evolutionary Robots," Series: Studies in Fuzziness and Soft Computing, Vol. 147, Springer-Verlag, Berlin/New York, ISBN: 3-540-20901-8, (2004). He has served as an *Organizing Chair, IEEE 2008 11th International Conference on Computer and Information Technology (ICCIT 2008) and Workshops, held during 24-27 December, 2008 at KUET*. Currently, he is working as a Technical Support Team Consultant for Bangladesh Research and Education Network (BdREN) in the Higher Education Quality Enhancement Project (HEQEP) of University Grants Commission (UGC) of Bangladesh.

FPGA Implementation of 5/3 Integer DWT for Image Compression

M.Puttaraju[1]
Professor, Department of Medical Electronics
DayanandaSagra College of Engineering
Bangalore, India

Dr.A.R.Aswatha[2]
Professor, Department of Telecommunication
DayanandaSagra College of Engineering
Bangalore, India

***Abstract*— The wavelet transform has emerged as a cutting edge technology, in the field of image compression. Wavelet-based coding provides substantial improvements in picture quality at higher compression ratios. In this paper an approach is proposed for the compression of an image using 5/3(lossless) Integer discrete wavelet transform (DWT) for Image Compression. The proposed architecture, based on new and fast lifting scheme approach for (5, 3) filter in DWT. Here an attempt is made to establish a Standard for a data compression algorithm applied to two-dimensional digital spatial image data from payload instruments.**

Keywords-2D-DWT; Lifting; CCDS; wavelet transform; 1DDWT.

I. INTRODUCTION

The wavelet transform has gained widespread acceptance in signal processing in general and in image compression research in particular. In applications such as still image compression, Discrete Wavelet Transform (DWT) based schemes have outperformed other coding schemes like the ones based on Discrete Cosine Transform (DCT). The DWT has been introduced as a highly efficient and flexible method for sub band decomposition of signals. The two dimensional DWT (2D-DWT) is nowadays established as a key operation in image processing. This is due to the fact that DWT supports features like progressive image transmission (by quality, by resolution), ease of compressed image manipulation, region of interest, etc. In addition to image compression, the DWT has important applications in many areas, such as computer graphics, numerical analysis, radar target distinguishing and so forth.

The transform coding part utilizes Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT) etc. Most widely used and efficient compression scheme JPEG uses Discrete Cosine Transform in its transform coding part. It splits the whole image into 8X8 pixel blocks and then for each block DCT is applied. The DCT has the disadvantage of blocking artefacts (blurring of images at edges of the image blocks) at higher compression ratios. This is overcome by the use of Discrete Wavelet Transform, since it considers image as a whole and moreover, it does not have pre-processing of image (splitting it into 8X8 pixel blocks). Thus for the same quality of the output image DWT has better compression ratios.

DWT has traditionally been implemented by convolution. Such an implementation demands both a large number of computations and a large storage—features that are not desirable for either high-speed or low-power applications. Recently, a lifting-based scheme that often requires far fewer computations has been proposed for the DWT [1].The transform coding part which we would be developing will be a 5/3 Integer Discrete Wavelet Transform (DWT). Suitable architecture is selected (lifting) [1] and the core is developed using VHDL.

II. CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS

The Consultative Committee for Space Data Systems (CCDS) is the recommendation for space data system standards. The main purpose of this paper is to establish a data compression algorithm applied to two-dimensional digital spatial image data from payload instruments based upon the recommendation [1].

III. LIFTING BASED DWT

The wavelet Lifting Scheme is a method for decomposing wavelet transforms into a set of stages. The convolution-based 1-D DWT requires both a large number of arithmetic computations and a large memory for storage. Such features are not desirable for either high speed or low-power image processing applications. The main feature of the lifting-based wavelet transform is to break-up the high pass and the low pass wavelet filters into a sequence of smaller filters. The lifting scheme requires fewer computations compared to the convolution-based DWT. Therefore the computational complexity is reduced to almost a half of those needed with a convolution approach [2][3].The Fig.1 illustrates Lifting Concept.

The main advantages of lifting scheme are as follows:

I) It allows a faster implementation of the wavelet transform.

II) The lifting scheme allows a fully in-place calculation of the wavelet transform. In other words, no auxiliary memory is needed and the original signal (image) can be replaced with its wavelet transform.

III) With the lifting scheme, the inverse wavelet transform can immediately be found by undoing the operations of the forward transform. In practice, this comes down to simply reversing the order of the operations and changing each + into a - and vice versa.

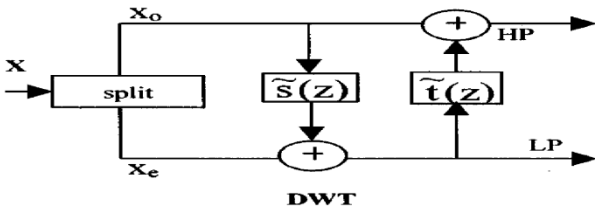


Fig. 1 Lifting Scheme

The three main steps of Lifting are

a) *Predict* step: where the odd samples are multiplied by the time domain equivalent of s(z) and are added to the even samples.

b) *Update* step: where updated even samples are multiplied by the time domain equivalent of t(z) and are added to the odd samples.

c) *Scaling* step: where the even samples are multiplied by 1/K and odd samples by K .

A. Two-Dimensional Discrete Wavelet Transform

The basic idea of 2-D architecture is similar to 1-Darchitecture. A 2-D DWT can be seen as a 1-D wavelet transform along the rows and then a 1-D wavelet transform along the columns, as illustrated in Figure 2.The 2- DWT operates in a straightforward manner by inserting array transposition between the two 1-D DWT. The rows of the array are processed first with only one level of decomposition. This essentially divides the array into two vertical halves, with the first half storing the average coefficients, while the second vertical half stores the detail coefficients. This process is repeated again with the columns, resulting in four sub-bands (see Fig.4) within the array defined by filter output. The LL sub-band represents an approximation of the original image, the LL1 sub-band can be considered as a2:1 sub-sampled (both horizontally and vertically) version of the original image.

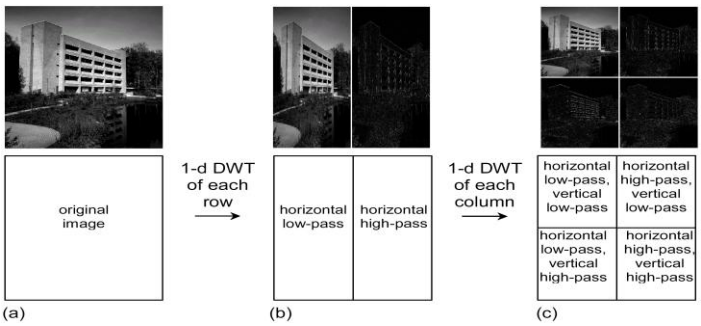


Fig .2 2_d DWT(one level)

The other three sub-bands HL1,LH1, and HH1 contain higher frequency detail information (mostly local discontinuities in the edges of the image). This process is repeated for as many levels of decomposition as are desired.

Here level of decomposition is chosen as three. To increase compression effectiveness, correlation remaining in the LL sub band after the 2-d DWT decomposition is exploited by applying further levels of DWT decomposition to produce a multi-level 2-d DWT (Fig.3). This produces the pyramidal decomposition [4].

CCSDS Recommendation standard specifies three level decomposition. As from the Figure.4 it is clear. Original Image undergoes first level decomposition to obtain four sub bands namely LL1, HL1, LH1, and HH1. Now the LL1 undergoes the second level decomposition to obtain the sub bands LL2, HL2, LH2 and HH2. In the final stage LL2 undergoes third level decomposition to obtain the sub bands LL3, HL3, LH3 and HH3. The DWT stage performs three levels of 2-D wavelet decomposition to obtain 10 sub bands as shown in the fig.4. Once LL1 undergoes first level decomposition remaining HL1, LH1 and HH1 are stored in DWT buffer as shown in fig 3.Similarly HL2, LH2 and HH2 are stored in buffer during next level decomposition. at each level of decomposition, the LL sub band from the previous level is decomposed, using a 2-d DWT, and is replaced with four new sub bands. Each new sub band is half the width and half the height of the LL sub band from which it was computed. Each additional level of decomposition thus increases the number of sub bands by three.

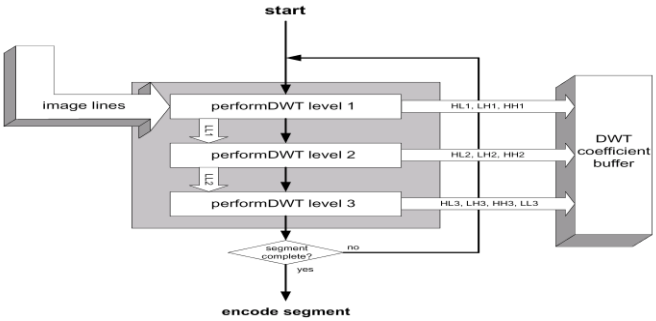


Fig.3 Program and Data Flow of DWT Module

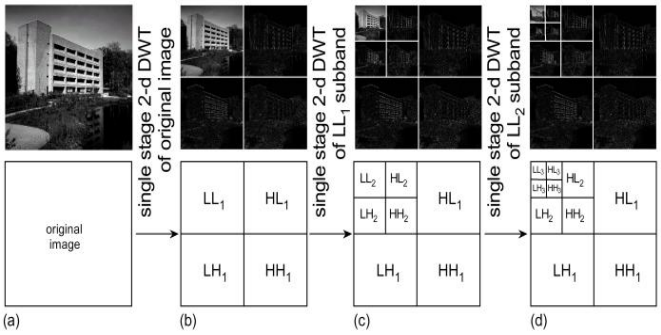


Fig. 4: Three-Level 2-d DWT Decomposition of a Image

IV. IMPLEMENTATION

CCSDS Recommended Standard use of a three-level, two dimensional (2-d), separable Discrete Wavelet Transform (DWT) with five and three taps for low- and high-pass filters, respectively.fig 5 shows the general schematic coder .

Two specific 1-d wavelets are specified with this Recommended Standard [1] Integer DWT which strictly supports lossless compression. Integer DWT is purely reversible i.e. reconstructed image will be same as original image.

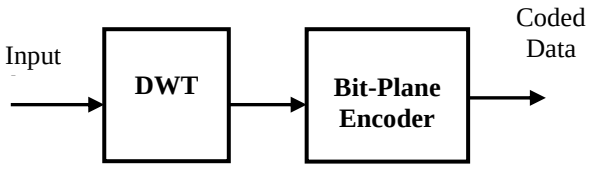


Fig .5 General Schematic of the coder

1-d Integer DWT maps a signal vector(shown in equation 1) to two sets of wavelet coefficients, one high-pass set, *Dj*, and one low pass set, Cj, in accordance with equations 2 and3[1] [2] [5]

For N>2 let

$\{X_0, X1, \ldots\ldots\ldots\ldots X_{2N-1}\}$ (1)

$$D_j = x_{2j+1} - \left\lfloor \frac{1}{2}(x_{2j}) + \frac{1}{2}(x_{2j+2}) \right\rfloor \quad (2)$$

For j=1…N-3

$$Cj = x_{2j} + \left\lfloor \frac{D_{2j-1} + D_{2j+1} + 2}{4} \right\rfloor \quad (3)$$

For j=1…N-1

Where " $\lfloor\ \rfloor$ " indicates floor brackets. Equations (2) and (3) define the integer transform that shall be used with this Recommended Standard. Given input values xi**,** the **Dj**values in equation (2) shall be computed first and used subsequently to compute **Cj**values in equation (3).These equations are "Lifting Equations".

A. Lifting Equations Implementation

The wavelet Lifting Scheme is a method for decomposing wavelet transforms into a set of stages .As compared to convolutional method the arithmetic computations required is less i.e computational complexity is reduced by half .Its implementation is described in fig.6

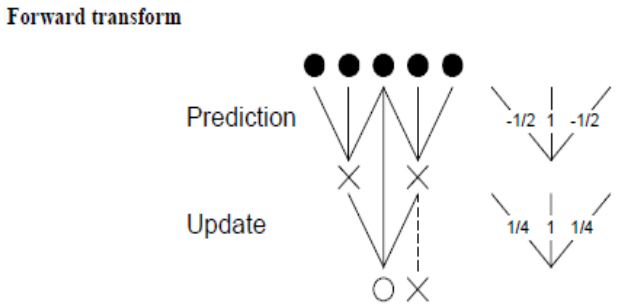


Fig .6 Lifting Concepts

The Figure .6 shows lifting concept where the odd samples are predicted from even samples which are called as prediction stage to obtain High Pass coefficients [6]. Then by using high Pass coefficients, Low Pass is updated which is the next step of lifting. Analysis of filter coefficients are shown in table no 1.

Table 1:Filter Coefficients for the 5/3 Integer DWT Filter

i	Low pass Filter hi	High pass filter gi
0	¾	½
±1	1/4	-1/4
±2	-1/8	

V. VHDL IMPLEMENTATION

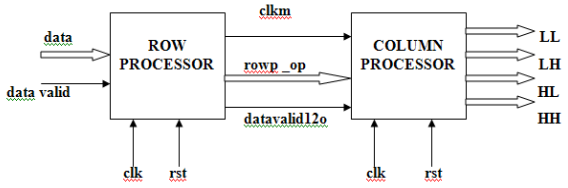


Fig .7.1 Block diagram for 1-level 2d-dwt

This is the basic block diagram for 1-level 2d-dwt which will produce the 1st level 4 sub bands ,among them only LL band is used for the 2nd level dwt[7].

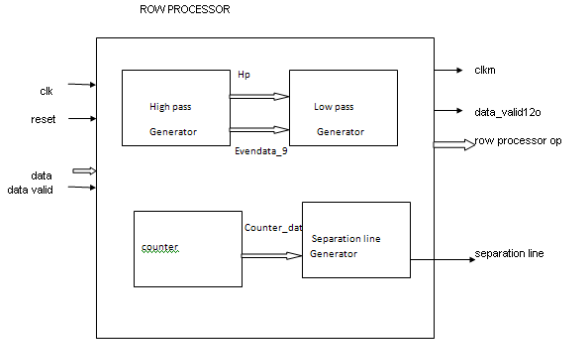


Fig 7.2 Block diagram of ROW processor

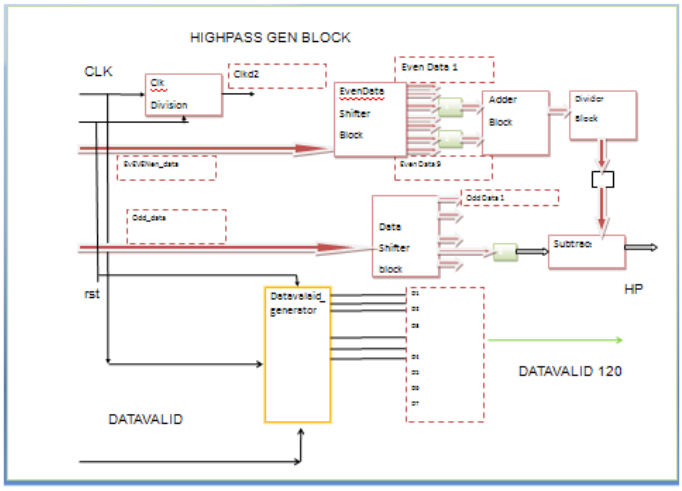


Fig 7.3 Block diagram of High pass generation block (ROW processor)

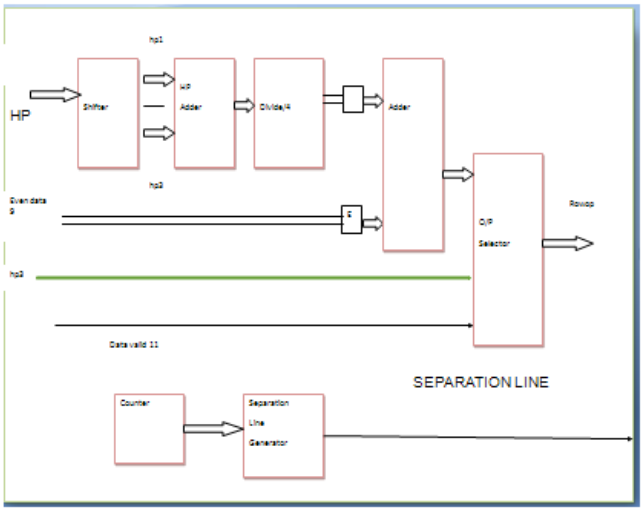


Fig 7.4 Block diagram of Low pass generation block (ROW processor)

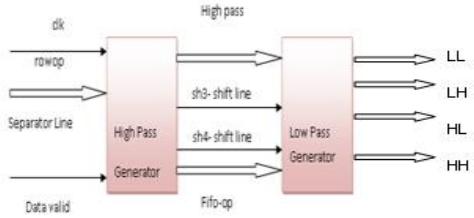


Fig 7.5 Block diagram of column processor

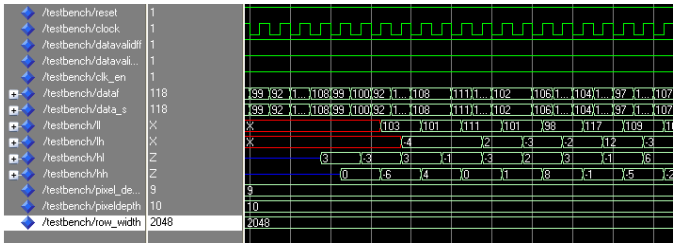


Fig 7.6 simulation results of 1 level 2d- DWT

VI. MEASURING OF PERFORMANCE

Mean Squared Error (MSE): MSE [1] is defined as follows:

$$\mathbf{MSE} = \frac{1}{M}\sum_{i=1}^{M}(x_i - \hat{x}_i)^2$$

Where, *M* is the number of elements in the signal, or image. For example, if we wanted to find the MSE between the reconstructed and the original image, then we would take the difference between the two images pixel by pixel, square the results, and average the results. Peak Signal to Noise Ratio (PSNR): The PSNR [1] is defined as follows:

$$\mathbf{PSNR} = 10\log_{10}\left(\frac{(2^n - 1)^2}{MSE}\right)$$

Where, *n* is the number of bits per symbol. As an example, if we want to find the PSNR between two 256 gray level images, then we set *n* to 8 bits. The following table no 3 give the psnr and rms calculation done for the 128×128 size of different images

Table 2: performance measures results for 1 level 2D DWT (with flooring and ceiling function)

Image	Avg error	RMS error	PSNR	SNR
Lenna	0.0	0.0	infinity	infinity
Baboon	0.0	0.0	infinity	infinity
Peppers	0.0	0.0	infinity	infinity
cameraman	0.0	0.0	infinity	infinity
goldhill	0.0	0.0	infinity	infinity

VII. RESULTS AND OBSERVATION

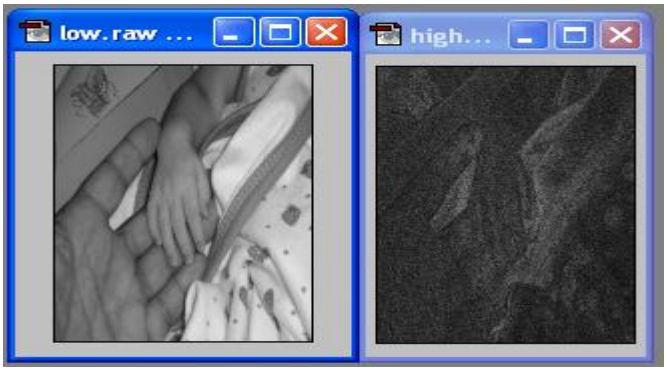


Fig.8.1 vhdl row processor output

Fig .8.2 column processor output.

Fig.9 Three –Level decomposition of Lena Image using C++ code.

The DWT level chosen is one for FPGA implementation in this paper.

VIII. CONCLUSION

In this paper, an approach is made proposed architecture for the5/3 Integer 2D-DWT to meet the requirements of real-time image processing.As mentioned in CCSDS document for three level decomposition of 9/7 Integer DWT is used to implement 5/3 filter.

The proposed architecture has been correctly verified by writing the code using VHDL Language. The code is synthesized using Axcelerator FPGA family.The estimated frequency of operation is around 60MHz.

REFERENCES

[1]. Image Data Compression. "Report Concerning Space Data System Standards", CCSDS 120.1-G-1. Green Book. Issue 1. Washington, D.C.: CCSDS, June 2007.

[2]. S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," IEEE transactions on Pattern Analysis and Machine Intelligence, vol. 11, no. 7, 1989, pp. 674–693.

[3]. W. Sweldens, "The Lifting Scheme: A Custom-Design Construction of Biorthogonal Wavelets", Applied and Computational Harmonic Analysis, Vol. 3, NO. 15, pp. 186-200, 1996.

[4]. K. Andra, C. Chakrabarti, and T. Acharya, "A VLSI architecture for lifting-based forward and inverse wavelet transform", IEEE Trans. Signal Processing, vol. 50, no. 4, pp. 966-977, April 2002.

[5]. A. R. Calderbank, I. Daubechies, W. Sweldens, and B.-L. Yeo, "Wavelet transforms that map integers to integers", Technical report, Deportment of Mathematics, Princeton University, 1996.

[6]. G. Knowles, "VLSI architecture for the discrete wavelet transform", Electronic Letters, vol. 26, no.5, pp. 1184-1185, July 1990.

[7]. Lossless Data Compression. Recommendation for Space Data Systems Standards. CCSDS 121.0-B-1. Blue Book. Issue 1. Washington D.C.: CCSDS, May 1997.

Request Analysis and Dynamic Queuing System for VANETs

Ajay Guleria
Department of Computer Centre
PU Chandigarh, India

Narottam Chand
Department of CSE
NIT Hamirpur, India

Mohinder Kumar
Department of CSE
NIT Hamirpur, India

Lalit Awasthi
Department of CSE
NIT Hamirpur, India

***Abstract*—Vehicular Ad hoc Network (VANET) is a kind of mobile ad hoc network using the capabilities of wireless communication for Vehicle-to-Vehicle and Vehicle-to-Roadside communication to provide safety and comfort to vehicles in transportation system. People in vehicles want to access data of their interest from Road Side Unit (RSU). RSU need to schedule these requests in a way to maximize the service ratio. In this paper we have proposed new methods for careful analysis of incoming requests to find whether these requests can be completed within deadline or not and to provide dynamic service queue. Simulation results show that the proposed schemes increase the service ratio significantly.**

Keywords- Road Side Uni; Service Ratio; Propagation Delay; Service Queue.

I. INTRODUCTION

Vehicular ad hoc network (VANET) is emerging as important technology to provide safety and comfort to vehicles in transportation system. It is special type of mobile ad hoc network with highly mobile nodes. Federal Communications Commission has allocated 5.850-5.925 GHz portion of the spectrum for inter-vehicle communication (IVC) and vehicle to roadside communication (VRC) [1, 2]. VANETs are not purely mobile ad hoc network; they have fixed points called Road Side Units (RSUs) to provide services to vehicles. RSU can provide safety, local news and advertisement, music, radio, video, etc. [3, 4, 5, 6]. Applications of VANET can be broadly classified into two categories.

A. *Safety related applications*: accident related alerts, red light warning, etc.

B. *Non-safety related applications:* these applications include downloading audio/video programs, digital map, internet related services, traffic information, weather forecast and other communication applications.

Some of the major challenges for communication in VANETs are high mobility, dynamically changing topology, sparsely located nodes and very short duration of communication. So, serving the requested data items to vehicles before it goes outside the coverage of RSU is very important. This paper makes the following contributions:

We have proposed an algorithm to check whether the incoming request can be completed before its deadline or not. Proposed an algorithm to provide dynamic service queuing to work efficiently under variable density of traffic.

Conducted simulation to evaluate the performance of VANETs when performing operations with both proposed algorithms.

Rest of the paper is organized as follows. Section II discusses related work in this field. Section III describes system model. Section IV proposes algorithms for request analysis and dynamic queuing system.Section V talks about simulation environment and results.Finally Section VI is devoted to concluding remarks.

II. RELATED WORK

Major research challenges in VANETs are introduced in [7, 8 and 9]. High mobility of vehicles is main challenge in VANETs which leads to short deadline to access data from RSU and causes highly dynamic topology. In case of vehicle to roadside data access there is more than one vehicle under coverage of one RSU. So multiple vehicles can send data upload/download request to RSU. Because deadlines are short therefore RSU needs to process these requests efficiently in terms of time. Many broadcasting algorithm have been proposed to reduce the waiting time [10, 11,12].

In [13] Acharya and Muthukrishnan proposed a data scheduling algorithm called longest total stretched first (LTSF) which is based on a new metric called stretch i.e. service ratio of the response time of a request to its service time. LTSF optimizes stretch and maintains balance between worst case and average case but implementation of LTSF for large system is not practical because server needs to recalculate stretch for every data item with pending request, to find next data to be broadcasted.

In [14] Xu et al. proposed online scheduling algorithm for time critical on demand data broadcast but they assumed that data can only be updated by server i.e. vehicle can only request download, it does not allow vehicles to update urgent data. Jiang and Vaidya in [15] proposed periodic push based broadcast, which is not well suited to VANET applications.

In [16] Zhang et al. proposed a vehicle platoon aware data access called V-PADA. In this scheme vehicles contribute their part of buffer to replicate data for others in the same platoon and share data with others. When vehicle leaves a platoon it prefetches interested data and transfers its buffered data to other vehicles in advance so that they can still access the data after it leaves.

V-PADA consists of two components: first a vehicle platooning protocol to identify platoon formation and platoon splitting by using stochastic time series analysis, second a data management to guide platoon members to replicate and fetchmost suitable data to achieve high availability and low control overhead.

In [17] Zhang et al. first proposed D*S algorithm, further optimized downloading by using D*S/N and optimized uploading by using D*S/R while maintaining different queues for download and upload requests. The algorithm assigns different bandwidth to these queues and serves upload requests on basis of service rate of data items in past.

None of the earlier data access schemes considered the optimization of service queue and incoming request analysis to remove the various sources of time wastage in data access scheme for VANETs. In contrast we have provided algorithms for analysis of incoming data and to make service queue size dynamic to increase the deadline by reducing the time wastage. The simulation results show that the proposed algorithms significantly improve the service ratio.

III. Ststem Model

In vehicular ad hoc networks there are two communicating entities i.e. vehicles and roadside unit (RSU). Each vehicle in VANET is equipped with On Board Unit (OBU) which has transceiver, computational power and omnidirectional antenna. RSU has transceiver, antenna, processor, sensors. RSU manages data and provides services to vehicle. Vehicles use services provided by RSU. Communication can be either inter-vehicle or vehicle-to-infrastructure. Fig. 1shows various steps involved in communication between vehicle and RSU.

Operation sequence of VANETs can be summarized in Fig. 2.

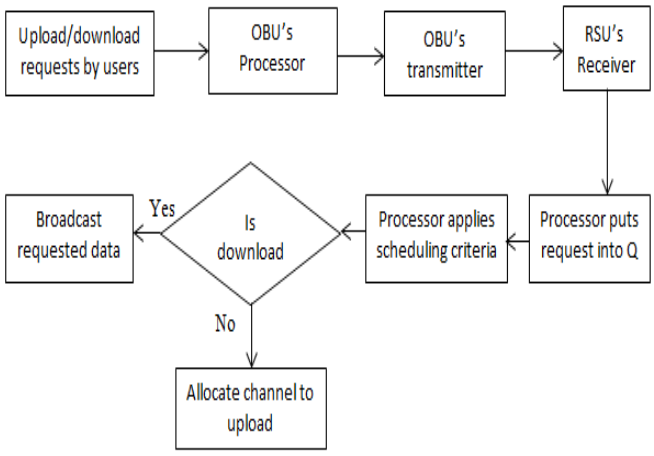


Figure 1.Various steps in communication between vehicleand RSU.

RSU is generally placed at intersections to increase service ratio and safety. We have also assumed that RSU is placed at intersection and vehicles are synchronized with Global Positioning System(GPS). Further, we have assumed two lanehighways for simplicity. The proposed scheme is applicable to multilane highways without any major change. OBU on vehicle is capable of calculating average speed of vehicle for last few minutes.

A request from vehicle is represented by five tuple as given below:

<VEHICLE_ID, DATA_ID, AVG_SPEED, CURRENT_LOCATION, OP_CODE>

VEHICLE_ID: Unique Identity of Vehicle.

DATA_ID: Unique Identification code of requested data.

AVG_SPEED: Average Speed of vehicle.

CURRENT_LOCATION: Current Location of vehicle.

OP_CODE: Operation Code i.e. either upload or download.

It is assumed that RSU maintains single queue for upload and download requests.RSU is capable of determining CURRENT_LOCATION. Hence it can find the direction of movement of vehicle.

IV. Proposed Schemes

In this section we have proposed two independent algorithms PROCESS_REQUEST to find whether incoming request can be completed before its deadline or not and SRBAQS to provide dynamic queuing system.

As shown in Fig. 2, our first proposed algorithm PROCESS_REQUEST uses following notations:

LNS: Lane from north to south.

LSN: Lane from south to north.

LWE: Lane from west to east.

LEW: Lane from east to west.

(XN, YN): Location of last coverage point of RSU in north.

(XS, YS): Location of last coverage point of RSU in south.

(XE, YE): Location of last coverage point of RSU in east.

(XW, YW): Location of last coverage point of RSU in west.

(XI, YI): Point of intersection of horizontal and vertical dividers.

(XR, YR): Reference point (discussed below).

(XV, YV): Current location of vehicle.

Size(i): Size of data item requested in current request.

t: transfer rate of data from RSU to vehicle.

ΔT: Average propagation delay of request from vehicle to RSU.

TTRANSFER: Time required for transferring requested data to vehicle.

TLIFE: Connection life time i.e. duration for which vehicle will remain in range of RSU.

Reference point: It is last coverage point in direction of movement of vehicle when vehicle requested the data. Even if vehicle takes a left or right turn, reference point will remain

same i.e. change in direction of movement after request has been sent to RSU does not make any effect on the proposed scheme.

PROCESS_REQUEST uses procedure REFERENCE_POINT to find the reference point. REFERENCE_POINT procedure is as below.

1. *REFERENCE_POINT*(X_V, Y_V)
2. If (X_V, Y_V) ∈ L_{NS}
3. then return (X_S, Y_S)
4. If (X_V, Y_V) ∈ L_{SN}
5. then return (X_N,Y_N)
6. If (X_V, Y_V) ∈ L_{EW}
7. then return (X_W, Y_W)
8. If (X_V, Y_V) ∈ L_{WE}
9. then return (X_E, Y_E)

In procedure REFERENCE_POINT lines 2 and 3 return reference point as (XS, YS) if vehicle is in lane from north to south i.e. moving from north to south direction. Lines 4 and 5 return reference point as (XN, YN) if vehicle is in lane from south to north i.e. moving from south to north direction. Lines 6 and 7 return reference point as (XW, YW) if vehicle is in lane from east to west i.e. moving from east to west direction. Similarly lines 8 and 9 return reference point as (XE, YE) if vehicle is in lane from west to east i.e. moving from west to east direction.

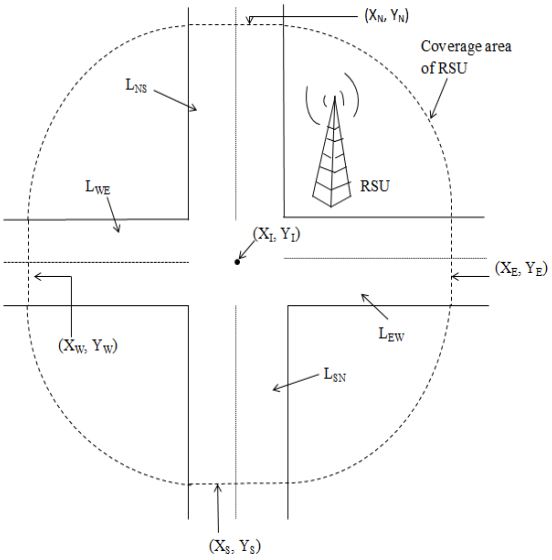


Figure 2. System model and notations.

1. PROCESS_REQUEST(XV, YV, AVG_SPEED)
2. (XR, YR)=REFERENCE_POINT(XV, YV)
3. d1 = √((XR-XV)2+(YR-YV)2)
4. if ((XR==XS)||(XR==XN))
5. d2 = √((XI-XW)2+(YI-YW)2)+ √((XI-XV)2+(YI-YV)2)
6. d3 = √((XI-XE)2+(YI-YE)2)+ √((XI-XV)2+(YI-YV)2)
7. if ((XR==XW)||(XR==XE))
8. d2 = √((XI-XN)2+(YI-YN)2)+ √((XI-XV)2+(YI-YV)2)
9. d3 = √((XI-XS)2+(YI-YS)2)+ √((XI-XV)2+(YI-YV)2)
10. d=min {d1, d2, d3}
11. di =d-(ΔT * AVG_SPEED)
12. TLIFE=di/AVG_SPEED
13. TTRANSFER=Size(i)/t
14. if (TLIFE >=TTRANSFER)
15. then put request in Queue
16. else
17. reject the request

Algorithm PROCESS_REQUEST at RSU determines whether the coming request can be completed within deadline or not. This algorithm takes current location of vehicle i.e. (XV, YV) and AVG_SPEED as input. Other values including Size(i) are known to RSU.

Line 2 calls the procedure REFERENCE_POINT on current location of vehicle to find the reference point. In line 2 (XA, YA) = (XB, YB) implies XA = XB and YA= YB. Line 3 computes distance between current location of vehicle and reference point, it is case when vehicle does not take any turn i.e. distance between (Xv, YV) and reference point computed in step 2 i.e. (XR, YR). Line 4 checks whether vehicle is moving in LNS or LSN i.e. vehicle can take turn to either west or eastside. If result of line 4 is true then line 5 computes total distance to be travelled by vehicle if it takes turn to west and line 6 computes total distance to be travelled by vehicle if it takes turn to east.

Line 7 checks whether vehicle is moving in LWE or LEW i.e. vehicle can take turn to either north or south side. If result of line 7 is true then line 8 computes total distance to be travelled by vehicle if it takes turn to north and line 9 computes total distance to be travelled by vehicle if it takes turn to south.

Line 10 finds the minimum distance which can be travelled by vehicle without going outside the coverage of RSU. Line 11 reduces distance to be travelled by vehicle, by the distance vehicle has travelled till its request reaches the RSU. Line 12 finds the time period for which vehicle will remain under coverage of RSU. Line 13 finds the time to be taken to transfer data to vehicle by dividing the requested data size by transfer rate. Line 14 checks whether data can be transferred to vehicle before it goes out coverage of RSU, if yes, line 15 puts this request in Queue else line 17 rejects the request.

Service Ratio Based Adaptive Queuing System (SRBAQS)

Traffic density varies on road during day and night, on working days and holidays also if there is jam on other routes. If size of service queue is kept static then, there can be following problems:

1. If service queue size at RSU is very large then it will take very long time to start the scheduling process, because RSU will wait for service queue to become full. Even when there is limit on waiting time, it will

affect the service ratio significantly as smaller deadlines are major constraint in VANETs.

2. If service queue size is very small it will start the scheduling process too early. It can reduce the service ratio e.g. let size of service queue is two and first two requests arrived are of very large size. Then it can block subsequent smaller size requests.

SRBAQS algorithm avoids this scenario by making service queue size dynamic depending on the service ratio. It uses following inputs:

λ: Service ratio of past x requests

x: number of requests processed in past

n: Service queue size

s: Number of requests successfully served, initially s=0

SC: Service ratio for current service cycle

λn: New service ratio after completion of current service cycle

SIZE(Q): Number of requests in queue

Success (requesti): Returns true if requesti has been completed successfully

1. SRBAQS (λ, x, n)
2. while(true)
3. while(SIZE(Q)!=n)
4. if new request has arrived put request in service queue Q
5. for i←0 to n do
6. process requesti according to scheduling criteria
7. if Success(requesti)=true
8. then s=s+1
9. SC=s/n
10. λn= (λx + SCn)/(x+n)
11. x=x+n
12. If (SC>λ)
13. then n=n+ n/2
14. else
15. n=n-n/2
16. λ=λn

NOTE: loops in SRBAQS are nested according to indentation.

Line 1of this algorithm shows that algorithm SRBAQS takes service ratio in past, number of requests processed in past and initial service queue size as input. Initial service queue size can be 1 or 2 whereas other two parameters are computed in past for very first service cycle these two can be any appropriate values. Line 2 of this algorithm shows that RSU will repeat same set of activities. Line 3 and 4 make service queue full before starting scheduling process. Lines from 5 to 8 process the service queue according to scheduling criteria and variable s counts the number of successful requests. Scheduling criteria in Line 6 can be

FCFS (First Come First Serve): Requests in service queue are served in sequence of their arrival.

EDF (Earliest Deadline First): Requests are served in increasing sequence of their deadline. Request having earliest deadline will be served first.

SDF (Smallest Data size First): Request having smallest data size will be served first and so on.

Lines 7 and 8 counts number of requests successfully served in current service cycle. Line 9 computes the service ratio of last service cycle. Line 10 computes the new service ratio. Line 11 computes total number of requests served including requests served in past. Lines 11 to 14 check if service ratio in last service cycle is greater than previous service ratio then increases service queue size by fifty percent else reduces the size of service queue by fifty per cent. Line 14 assigns new service ratio to current service ratio. After each iteration of while loop in line 2 s is initialized to zero.

Theorem: the proposed algorithm PROCESS_REQUEST computes the minimum distance to be travelled by the vehicle. Proof: Let the vehicle is moving on LWE and has requested some data then there can be two cases:

Case 1: vehicle is at some position from which it can take either a turn or move straight as shown in Fig. 3(a).

Case 2: vehicle cannot take any turn i.e. it moves only straight as shown in Fig. 3(b).

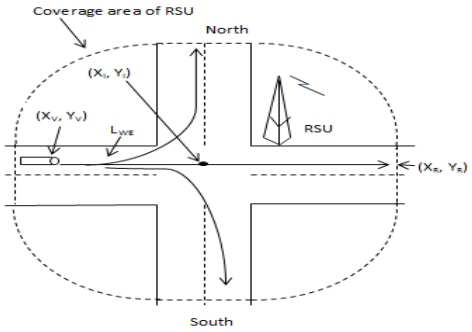


Figure 3(a). Case 1 when vehicle can take turn.

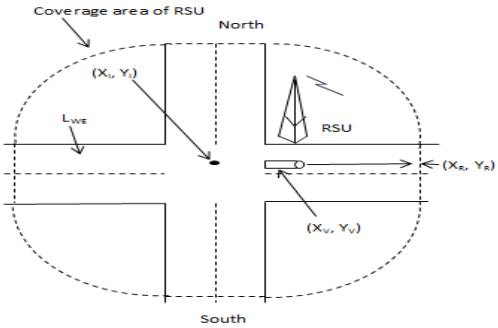


Figure 3(b). Case 2 when vehicle cannot take turn.

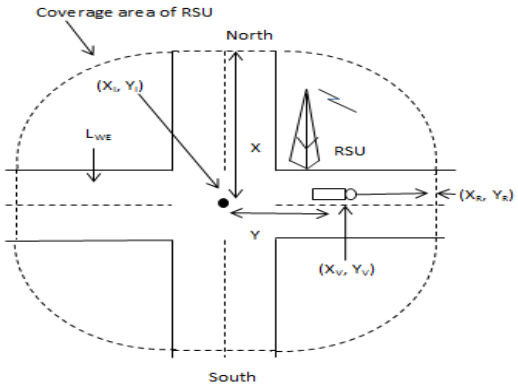


Figure 3(c). Path of d2.

In case 1, PROCESS_REQUEST computes distances d1, d2, d3 where:d1 is distance to be travelled by vehicle if it moves straight, d2 is distance to be travelled by vehicle if ittakes a turn to north, d3 is distance to be travelled by vehicle if it takes turn to south.

Proposed algorithm choses minimum of d1, d2 and d3. Hence distance to be travelled by vehicle is minimum distance.

In case 2, the proposed algorithm computes d1, d2, d3 although vehicle can move only straight.

$d2 = \sqrt{((XI-XN)2 + (YI-YN)2)} + \sqrt{((XI-XV)2+(YI-YV)2)}$

where $\sqrt{((XI-XV)2+(YI-YV)2)}$ is distance from vehicle to point of intersection. Let it is Y. See Fig. 3(c).

And $\sqrt{((XI-XN)2+(YI-YN)2)}$ is distance from point of intersection to north direction let it is X. See Fig. 3(c).

So, d2=X+Y

even for very small value of Y

$d2 < d1$ (1)

d2 can be greater than d1 iff RSU is placed at some place other than intersection.

Similarly, $d3<d1$ (2)

From Equation (1) and (2), we can say that d1 is minimum distance to be travelled by vehicle.

V. SIMULATIONENVIRONMENT AND RESULTS

We have simulated the proposed algorithms i.e. PROCESS_REQUEST and SRBAQS using Dev-C++ 4.9.9.2. The system model has been implemented by appropriately and randomly generating request ids, data ids, operation codes, data size, deadlines and other parameters. We have tested the performance of both algorithms over three data scheduling algorithms discussed earlier i.e. FCFS (First Come First Serve), EDF(Earliest Deadline First), SDF(Smallest Data size First). For evaluation of these algorithms, we have used Service ratio as performance metric i.e. ratio of number of requests served successfully to total number of requests. In each graph Y axis denotes the service ratio and X axis number of times window i.e. service queue processed.

Fig.4 show the performance of FCFS before and after implementing PROCESS REQUEST. Fig.5 and 6 show the performance of EDF and SDF respectively before and after implementing PROCESS_REQUEST.

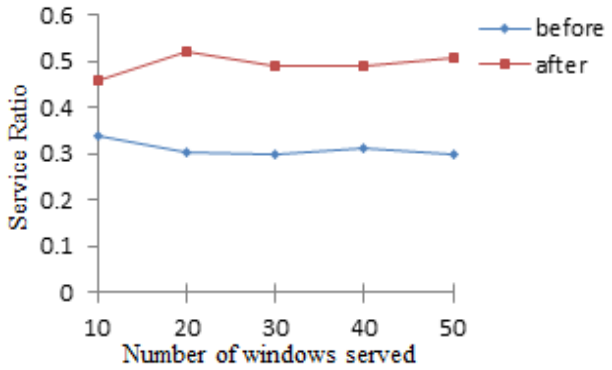


Figure 4. Performance of FCFS before and before and after implementing PROCESS_REQUEST.

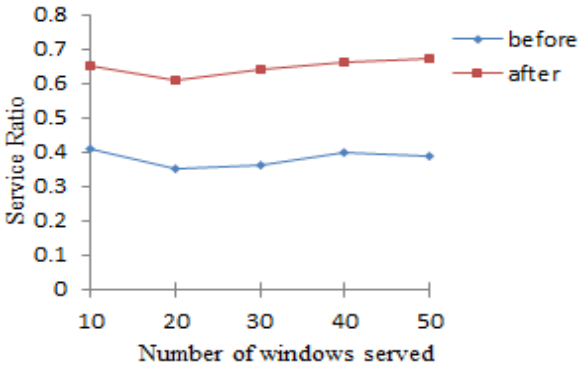


Figure 5. Performance of EDF before and after implementing PROCESS_REQUEST.

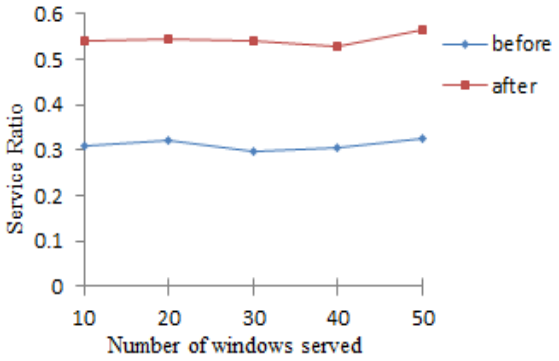


Figure 6. Performance of SDF before and after implementing PROCESS_REQUEST.

Fig. 4, 5 and 6 show that the PROCESS_REQUEST algorithm improves the performance of scheduling algorithms significantly. This improvement is achieved because the PROCESS_REQUEST eliminates the data access requests which cannot be completed within their deadline therefore avoiding the wastage of time on unnecessary processing time of these requests.

Fig. 7 shows the performance of FCFS before and after implementing both PROCESS_REQUEST and SRBAQS.

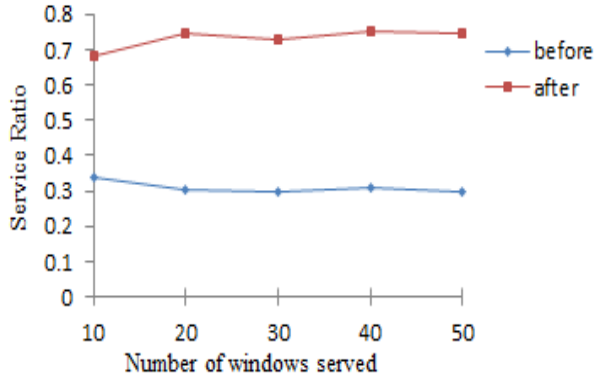


Figure 7. Performance of FCFS before and after implementing both PROCESS_REQUEST and SRBAQS.

Fig. 8 and 9 show the performance of EDF and SDF before and after implementing both PROCESS_REQUEST and SRBAQS.

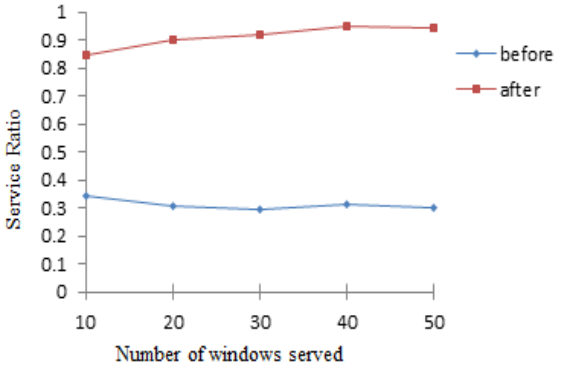


Figure 8. Performance of EDF before and after implementing both.

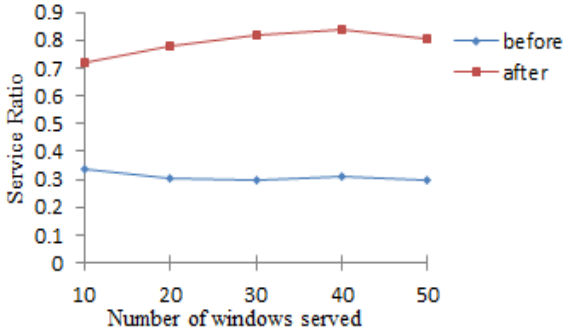


Figure 9.Performance of SDF before and after implementing both PROCESS_REQUEST and SRBAQS.

Fig. 7, 8 and 9 show that implementation of both PROCESS_REQUEST and SRBAQSfurther improve the performance. This performance gain is because PROCESS_REQUEST avoid the processing of unnecessary requests and SRBAQS make the queue size dynamic. Hence simulation results show that PROCESS_REQUEST and SRBAQS significantly improve the performance of data scheduling algorithms.

VI. Conclusion

In this paper we have proposed two algorithms PROCESS_REQUEST and SRBAQS. PROCESS_REQUEST checks whether the incoming request can be served before its deadline or not. If request cannot be completed within its deadline then it is not inserted in service queue hence rejected. It improves the processing time of other requests and service ratio. Static queue size causes various problems such as long waiting time if traffic density is very low and poor performance of scheduling algorithm if traffic density is very high. SRBAQS solves both the problems by providing dynamic queue size i.e. queue size varies depending upon service ratio. Simulation results show that both algorithms improve the performance of data scheduling algorithms FCFS, EDF and SDF.

In future, we will take other issues into consideration on scheduling such as different types of data items, multiple queues, etc. Further, other unique challenges in VANETs such as routing, clustering, data caching will motivate further research in this area.

References

[1] IEEE Standard for Information Technology-Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements," IEEE Std. 802.11e-2005(Amendment to IEEE Std. 802.11, 1999 Edition (Reaff 2003), 2005.

[2] "IEEE 802.11p D3.0," in IEEE StandardActivitiesDepartment, July 2007.

[3] J. Yin, T. Eibatt, G. Yeung, B. Ryu, S. Habermas, H. Krishnan, and T. Talty, "Performance Evaluation of Safety Applications over DSRC Vehicular Ad hoc Networks,"in Proc. of the 1st ACM int. workshop on Vehicular ad hoc network, Oct. 1-1, 2004, pp.1-9.

[4] X. Yang, J. Liu, F. Zhao, and N. Vaidya, "A Vehicle-to-Vehicle Communication Protocol for Cooperative CollisionWarning," in Proc. of 1st Int. Conf. on Mobile and Ubiquitous Systems: Networking and Services, Aug. 22-26, 2004, pp. 114-123.

[5] T.Mak, K. Laberteaux, and R. Sengupta, "Multi ChannelVANET providing concurrent Safety and Commercial Services," in Proc. of the 2nd ACM int. workshop on Vehicular ad hoc networks, Sep. 02-02, 2005, pp. 1-9.

[6] S. B. Lee, G. Pan, J.S. Park, M. Gerla, and S. Lu,"Secure Incentives for Commercial Ad Dissemination in Vehicular Networks," inProc. of the 8th ACM int. symp. on Mobile ad hoc networking and computing, Sept. 9-14, 2007, pp. 150-159.

[7] K. Daniel, Wong, K. Tepe, W. Chen and M. Gerla, "Inter-Vehicular Communications,"in IEEE Wireless Communications, Vol.13, No. 5, pp. 6-7, Oct. 2006.

[8] R. A. Santos, A. Edwards and O. Alvarez, "Towards an Inter vehicle Communication Algorithm," The 3rd International Conference on Electrical and Electronics Engineering, September 6-8, 2006, pp. 1-4.

[9] M. Durresi, A. Durresi and L. Barolli, "Adaptive Inter Vehicle Communications," International Journal of Wireless Information Networks,Vol. 13, No. 2, pp. 151-160, April 2006.

[10] C. Su and L. Tassiulas, "Broadcast scheduling for information distribution," in Proc. of 16TH International Conference on Computer Communications, April 7-12, 1997, Vol. 1, pp. 109-117.

[11] D. Aksoy and M. Franklin, "R*w: a scheduling approach for large scale on-demand data broadcast," IEEE/ACM Transactions on Networking, Vol. 7, No. 6, pp. 846–860, Dec. 1999.

[12] R .Gandhi, S. Khuller, Y. Kim and Y. Wan, "Algorithms for minimizing response time in broadcast Scheduling," ALGORITHMICA, Vol. 38, No.4, pp. 597–608, 2004.

[13] S. Acharya and S. Muthukrishnan, "Scheduling On DemandBroadcasts: New Metrics and Algorithms,"in Proc. of the 4th annual ACM/IEEE int.conf. on Mobile computing and networking, Oct. 25-30, 1998, pp. 43-54.

[14] J. Xu, X. Tang and W. Lee, "Time-critical On-demand Data Broadcast: Algorithms, Analysis, and Performance evaluation," IEEE Transaction on Parallel Distributed Systems, Vol. 17, No. 1, pp. 3–14, 2006.

[15] S. Jiang and N. Vaidya, "Scheduling data broadcast to impatientusers," inProc. of the 1st ACM int. workshop on Data engineering for wireless and mobile access, Aug. 20-20, 1999, pp. 52–59.

[16] Y. Zhang and G. Cao, "V-PADA: Vehicle Platoon Aware Data Access in VANETs," IEEE transactions on vehicular technology, Vol.60, No. 5, pp. 2326-2339, June 2011.

[17] Y. Zhang, J. Zhao and G. Cao, "Service Scheduling of Vehicle-Roadside Data Access," in international journal of Mobile Networks and Application.,Vol. 15, No. 1, pp. 83-96, Feb. 2010.